Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaires
le 17/03/2021

Par : **Romain Birsal**

Titre : **Sélection des sinistres en cours et provisionnement des tardifs : Analyse de la performance des méthodes usuelles**

Confidentialité : ☒ NON ☐ OUI (Durée : ☐ 1 an ☐ 2 ans)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membres présents du jury de la filière

Entreprise : Prim'Act
Nom : Killian PAYEN
Signature :

Membres présents du jury de l'Institut des Actuaires

Directeur du mémoire en entreprise :

Nom : Frédéric PLANCHET
Signature :

Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels *(après expiration de l'éventuel délai de confidentialité)*
Signature du responsable entreprise

Secrétariat :

Signature du candidat

Bibliothèque :

Ecole Nationale de la Statistique et de l'Administration Economique (ENSAE)
5, avenue Henry Le Chatelier - 91120 PALAISEAU, FRANCE

Résumé

L'objectif de ce mémoire est de tester et analyser les méthodes de provisionnement classiques utilisées dans le cadre de l'arrêt de travail et plus particulièrement dans le cadre de l'incapacité temporaire de travail (ITT).

Pour cela, nous nous basons d'abord sur une structure par âge qui représente la population des actifs français. Cette structure est construite à partir des données de l'INSEE.
À partir de cette population, nous générons les entrées en arrêt de travail de chaque individu, en fonction du sexe et de l'âge, grâce aux taux fournis par la table d'incidence de Prim'Act.

Pour chaque individu en arrêt de travail, nous simulons la date de survenance du sinistre, la date de déclaration du sinistre ainsi que la durée de maintien en état d'incapacité de l'individu grâce à la table de maintien du BCAC (Bureau Commun des Assurances Collectives). Nous disposons ainsi, d'une réalité simulée comportant un nombre d'individus rentrés en incapacité au cours d'une année N avec connaissance de leur durée totale d'incapacité.

À partir de cette réalité, nous nous plaçons du point de vue de l'assureur qui cherche à provisionner, au 31/12/N, les sinistres survenus cette année. L'assureur, à cette date, n'a connaissance que des sinistres survenus et connus ainsi que leur ancienneté. Son objectif est donc d'estimer la charge des durées résiduelles d'arrêt de travail des sinistres en cours et également la charge des sinistres tardifs ou IBNR (Incurred But Not Reported).

Pour cela, nous utilisons différentes méthodes de provisionnement : une méthode ligne à ligne pour provisionner les sinistres en cours ; une méthode agrégée et une méthode basée sur la loi d'incidence pour estimer les sinistres tardifs. Ces méthodes nous donnent une estimation de la charge ultime des durées d'arrêt de l'année N que nous comparons avec les durées simulées. Il s'agit ensuite de répéter cette opération sur 500 simulations, afin d'obtenir une distribution empirique de la charge ultime estimée.

Enfin, nous discutons de la stabilité des méthodes de provisionnement lorsqu'il y'a une incertitude de modélisation lors de la construction des tables utilisées pour provisionner ainsi que lorsque le portefeuille présente un manque d'information au niveau individuel.

Mots-clés : *incapacité temporaire de travail*, *table d'incidence*, *table de maintien*, *simulation*, *provisionnement*, *IBNR*.

Abstract

The objective of this dissertation is to test and analyse the classic provisioning methods used in the context of work stoppages and more particularly in the context of temporary work disability (TWD).

In order to do this, we first base ourselves on the age structure of a population, which represents the French working population. This structure is constructed from INSEE data.
From this population, we generate the number of work stoppages according to the sex and age thanks to the rates provided by the Prim'Act incidence table.

For each individual in a work stoppage, we simulate the date the claim arises, the reporting date of the claim and the duration of the individual's incapacity thanks to the maintenance law of the BCAC (Bureau Commun des Assurances Collectives). We thus have a simulated reality with a number of individuals in incapacity to work during a given year N with the knowledge of their total duration of incapacity.

Given this situation, we take the point of view of the insurer who seeks to estimate, on 31/12/N, the claims that have occurred this year. The insurer, to this date is only aware of claims that have occurred and known as well as their seniority. His objective is therefore to estimate the burden of the residual duration of work stoppage of ongoing claims and the cost of late claims or IBNR (Incurred But Not Reported).

For this, we use different provisioning methods : a line-to-line method to provision ongoing claims; an aggregate method and a method based on the law of incidence to estimate IBNR.

These methods give us an estimate of the ultimate load of work stoppages duration of the year N. Then we compare it with the simulated durations. It is finally a question of repeating this operation on 500 simulations, in order to obtain an empirical distribution of the estimated ultimate charge.
Finally, we discuss the stability of these methods when there is modeling uncertainty due to issues such as table construction and a lack of indivual information in the portfolio.

Keywords : *temporary work disability*; *Law of incidence*, *Law of maintain*, *Simulation*, *provisioning methods*, *IBNR*.

Note de Synthèse

Contexte et problématique de l'étude

L'assurance est un secteur économique singulier qui est caractérisé par une inversion de son cycle de production. En effet, l'assureur reçoit de la part de son assuré une prime et s'engage, en contrepartie, à lui verser une prestation en cas de réalisation d'un risque aléatoire prévu dans le contrat pendant la période de couverture.

Ainsi, le principal enjeu pour une compagnie d'assurance est d'estimer le montant des réserves qu'elle aura à constituer afin d'honorer ses engagements en cas de réalisation du sinistre prévu par la police pendant la période de couverture. Les provisions sont définies comme l'espérance de la somme des flux futurs actualisés ajoutée avec la marge pour risque. Elles représentent une part très importante du passif du bilan de la compagnie d'assurance. Elles nécessitent donc une évaluation très précise.

L'objectif de ce mémoire est d'analyser la performance des différentes méthodes de provisionnement couramment utilisées sur le marché dans le cadre de l'arrêt de travail. Le cadre de l'étude désigne plus particulièrement l'incapacité temporaire de travail (ITT), c'est à dire le passage pour un individu de l'état actif à l'état d'incapacité en considérant sa durée de maintien dans cet état. La durée maximale de l'incapacité est de 3 ans, soit 36 mois.
Le but de l'assureur est donc d'estimer la durée résiduelle d'un individu en arrêt de travail afin de l'indemniser.

Pour cela, l'assureur peut utiliser différentes méthodes de provisionnement. Les méthodes généralement utilisées dans le cadre de l'arrêt de travail sont les méthodes de provisionnement ligne à ligne et les méthodes de provisionnement agrégées à base de triangles de paiements. Chacune présente des avantages et des inconvénients.

L'objectif de ce mémoire est d'analyser la performance de ces méthodes en confrontant l'estimation des provisions avec une simulation de la réalité qui représente au mieux la sinistralité d'arrêt de travail d'une population d'actifs français. Une méthode est qualifiée de plus précise qu'une autre si l'écart entre la provision et la simulation est plus faible que l'écart obtenu avec l'autre méthode.

Méthodologie

Le processus mis en place dans ce mémoire suit le plan suivant :

- Construire une structure par âge et par sexe qui représente la population d'actifs français ;
- Simuler à partir de cette structure la sinistralité d'incapacité de chaque individu : date de survenance, date de déclaration, durée de maintien en incapacité ;
- Provisionner en se plaçant du point de vue de l'assureur au $31/12/N$ les sinistres en cours et les tardifs à l'aide des différentes méthodes ;
- Comparer l'écart entre l'estimation des provisions et les durées réelles de la simulation ;
- Répeter l'opération sur 500 simulations pour comparer l'espérance et la volatilité empirique des provisions avec les durées simulées

Nous construisons d'abord une structure par âge et par sexe en nous basant sur les données de l'INSEE pour représenter la population des actifs français. La pyramide des âges (en milliers) suivante représente l'allure de la structure.

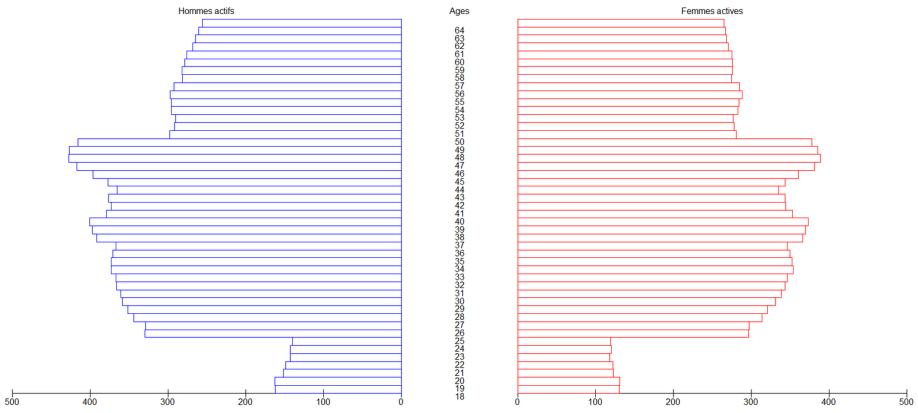


FIGURE 1 : Pyramide des âges de la population fictive des actifs français

A partir de cette structure, nous simulons les entrées en arrêt de travail à l'aide de la loi d'incidence donnée par la table d'expérience de Prim'Act. La table est construite pour de l'assurance emprunteur. Les franchises de ce type de contrat sont généralement de 3 mois, il est donc important de noter que les entrées en incapacité sont considérées avec 3 mois d'ancienneté.

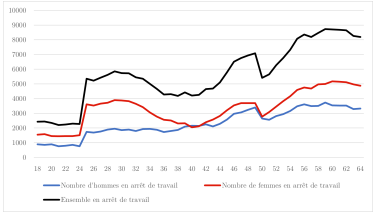
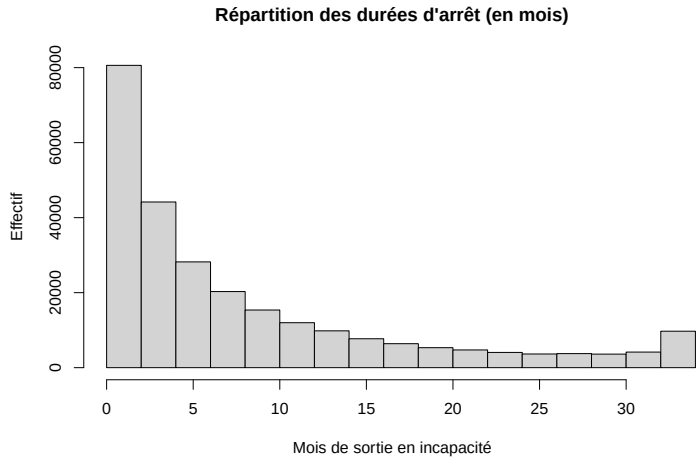


FIGURE 2 : Ensemble des individus (en milliers) entrés en arrêt de travail (avec 3 mois d'ancienneté)

Nous possédons ainsi une base d'individus entrés en arrêt de travail avec une ancienneté de 3 mois. A partir de cette base, nous simulons pour chaque individu les caractéristiques de la sinistralité c'est à dire la date de survenance, la date de déclaration ainsi que la durée de maintien en incapacité.

La date de survenance des sinistres est simulée avec une loi uniforme sur une année. Le délai de déclaration est simulé avec une loi log-normale dont la moyenne est de 2 mois. Enfin, la durée de maintien en incapacité est simulée à l'aide de la loi de maintien donnée par la table du BCAC (Bureau Commun d'Assurances Collectives).

Chaque individu possède donc en fonction de son âge une durée d'incapacité simulée. Il est important de noter que dans la suite du mémoire nous nous interessons à la durée indemnisée. De ce fait, les individus ont une durée d'indemnisation qui varie de 0 à 33 mois car ils ont 3 mois d'ancienneté.



Dans cette partie, nous nous plaçons du point de vue d'un assureur qui cherche à provisionner ses sinistres d'arrêt de travail au $31/12/N$. L'assureur quant à lui, n'a pas connaissance des durées totales précédentes. Il possède donc une base qui a été construite à partir de la simulation dont les durées d'arrêt sont censurées au $31/12/N$ et qui comprend uniquement les sinistres déclarés avant le $31/12/N$ en fonction du délai de déclaration.

Cette base contient donc 2 types de sinistres :

- les sinistres dont la durée d'indemnisation s'est terminée avant le $31/12/N$ que l'on appelle sinistres clos et dont on connait donc la prestation ;
- les sinistres dont l'individu est toujours en état d'incapacité au $31/12/N$ que l'on appelle sinistres en cours et dont la prestation reste à estimer.

Le but de cet assureur est donc de provisionner les sinistres en cours ainsi que d'estimer la charge des sinistres tardifs dont l'assureur n'a pas encore connaissance au $31/12/N$. Les charges des sinistres dans la suite de ce mémoire sont présentés en nombre de mois d'indemnisation.

Pour cela, nous utilisons 3 méthodes de provisionnement :

- une méthode de provisionnement ligne à ligne pour estimer la charge des sinistres en cours ;
- une méthode de provisionnement agrégée : Chain Ladder sur un triangle en nombre de sinistres pour estimer la charge ultime de l'année N et les tardifs ;
- une méthode de provisionnement basée sur la loi d'incidence pour estimer les tardifs.

L'approche ligne à ligne est basée sur le calcul d'un coefficient de provisionnement en mois d'indemnisation en fonction de l'ancienneté et l'âge de chaque individu.

La méthode Chain Ladder consiste en la construction d'un triangle de paiement en nombre de sinistres grâce à un historique sur 3 années généré de la même manière que la simulation précédente. Le nombre ultime de sinistres a ensuite été multiplié par un coût moyen de $7,62$ reflétant l'historique pour obtenir une estimation de la charge.

La dernière méthode de provisionnement est basée sur la loi d'incidence afin d'estimer le nombre de sinistres tardifs. Puis nous supposons que chaque tardif a une ancienneté de 2 mois d'indemnisation car le délai de déclaration des sinistres est d'en moyenne 2 mois. Nous réalisons ensuite une approche ligne à ligne sur ces derniers.

Nous avons d'abord testé ces méthodes sur un état du monde, c'est à dire sur une simulation de la réalité puis nous avons répété le processus décrit sur un échantillon de 500 simulations pour obtenir une distribution empirique de l'estimation de la charge ultime et des tardifs.

Résultats sur un état du monde

Les résultats obtenus sur un état du monde sont les suivants* :

Charge estimée (hors IBNR)	Charge simulée (hors IBNR)	Ecart relatif
1 655 864	1 651 673	0,25%

FIGURE 3 : Résultats de l'estimation des sinistres en cours grâce à la méthode ligne à ligne

Nombre de sinistres ultime estimé	Nombre de sinistres simulé	Ecart relatif
270 051	262 758	2,78%
Charge ultime estimée	Charge totale simulée	Ecart relatif
2 060 153	2 002 932	2,85%

FIGURE 4 : Résultats de l'estimation de la charge ultime grâce à la méthode de Chain Ladder

L'estimation des sinistres en cours à l'aide de la méthode ligne à ligne est assez précise avec un écart de $0,25\%$. Cette méthode reste précise car le risque est individualisé en fonction des caractéristiques des individus.
La méthode de Chain Ladder donne une estimation de la charge ultime assez correcte avec un écart de $2,7\%$

Estimation des IBNR

Méthode Chain Ladder		
Nombre d'IBNR estimé	Nombre d'IBNR simulé	Ecart relatif
52 718	45 425	16,05%
Charge d'IBNR estimée	Charge d'IBNR simulée	Ecart relatif
402 174	346 215	16,16%
Méthode basée sur la loi d'incidence		
Nombre d'IBNR estimé	Nombre d'IBNR simulé	Ecart relatif
46 517	45 425	2,40%
Charge d'IBNR estimée	Charge d'IBNR simulée	Ecart relatif
427 764	346 215	23,55%

FIGURE 5 : Résultats de l'estimation des IBNR

La méthode de Chain Ladder surestime le nombre d'IBNR avec un écart de $16,05\%$ tandis que la méthode basée sur la loi d'incidence donne une estimation assez correcte avec un écart de $2,40\%$. En ce qui conserne la charge, la méthode de Chain Ladder est cependant plus précise. En effet, l'hypothèse assez forte introduite sur l'ancienneté des tardifs dans la 2ème méthode a eu pour effet une surestimation assez importante de la charge des IBNR avec un écart de $23,55\%$

Résultats sur un échantillon de 500 simulations

Les résultats obtenus sur 500 simulations sont les suivants :

	Charge estimée (hors IBNR)	Charge simulée (hors IBNR)
Espérance empirique	1 660 071,77	1 664 256,56
Volatilité empirique	5 054,41	5 648,72

FIGURE 6 : Résultats de la méthode ligne à ligne sur 500 simulations

	Charge totale estimée	Charge totale simulée
Espérance empirique	2 047 518,70	2 014 241,00
Volatilité empirique	25 716,05	6 269,44

FIGURE 7 : Résultats de la méthode de Chain Ladder sur 500 simulations

La méthode ligne à ligne reste assez précise et peu volatile. La méthode de Chain Ladder reste assez précise aussi pour estimer la charge totale mais est beaucoup plus volatile à cause d'une irrégularité au sein de l'historique d'année en année lors des simulations.

Estimation des IBNR

	Charge des sinistres tardifs
Espérance empirique de la simulation	349 984,43
Charge des tardifs estimée avec Chain Ladder	
Espérance empirique	383 438,40
Volatilité empirique	24 982,65
Ecart relatif à la simulation	9,56%
Charge des tardifs estimée avec la loi d'incidence	
Espérance empirique	421 542,37
Volatilité empirique	4 355,39
Ecart relatif à la simulation	20,45%

FIGURE 8 : Résultats des méthodes sur l'évaluation des IBNR

Les deux méthodes ont tendance à surestimer la charge des tardifs. La méthode de Chain Ladder est beaucoup plus volatile (5 fois plus) que le l'autre méthode. En effet, puisque l'autre méthode est basée sur une approche ligne à ligne, elle reste assez stable empiriquement. Cependant elle reste peu précise à cause de l'hypothèse introduite.

Nous avons ensuite confronté les méthodes de provisionnement lorsque l'assureur commet une erreur d'estimation lors de la construction de ses tables d'expérience ainsi que lorsque le portefeuile de l'assureur présente un manque d'information individuelle.
Pour cela des chocs ont été introduits sur les tables d'entrée et de maintien utilisées pour provisionner. Nous obtenons les résultats suivants :

Impact des chocs sur la table d'entrée en incapacité

Chocs	Nombre d'IBNR estimé	Nombre d'IBNR simulé	Ecart relatif
10,00%	72 901	45 425	60,49%
5,00%	59 709	45 425	31,45%
2,00%	51 792	45 425	14,02%
1,00%	49 156	45 425	8,21%
0,50%	47 841	45 425	5,32%

Impact des chocs sur la table de maintien en incapacité

Chocs	Charge estimée	Charge simulée	Ecart relatif
20,00%	1 817 672	1 656 717	9,72%
15,00%	1 776 112	1 656 717	7,21%
10,00%	1 734 553	1 656 717	4,70%
5,00%	1 692 994	1 656 717	2,19%

Nous remarquons que les chocs effectués sur la table d'entrée entrainent un écart assez important sur l'estimation des tardifs. Les chocs effectués sur la table de maintien entrainent des écarts moins importants. Les ordres de grandeur des chocs introduits ont été comparés avec ce que l'on trouve en pratique lors des constructions des tables d'entrée et de maintien dans les différents mémoires. Ces ordres de grandeur semblent cohérents.

Enfin, nous avons étudié l'impact d'une connaissance imparfaite du portefeuille de l'assureur sur le provisionnement des sinistres en cours. Pour cela nous avons considéré que la base de l'assureur manque d'information au niveau individuel. La base contient les sinistres par tranches d'âge, un effectif par tranches d'âge ainsi qu'une ancienneté moyenne par tranches d'âge.

Afin de provisionner, l'assureur utilise des hypothèses basées sur un individu moyen puis utilise la méthode ligne à ligne sur celui-ci. Les hypothèses sont choisies en fonction de l'historique des années $N-2$ et $N-1$ et le provisionnement est effectué à l'aide de la table de maintien du BCAC ainsi que les tables de maintien d'expérience de la partie précédente.
Nous obtenons les résultats suivants :

Résultats du provisionnement avec la table du BCAC

Charge estimée	Charge simulée	Ecart relatif
1 711 749	1 656 717	3,32%

Résultats du provisionnement avec les tables d'expérience

Chocs	Charge estimée	Charge simulée	Ecart relatif
5,00%	1 756 324	1 656 717	6,01%
10,00%	1 800 899	1 656 717	8,70%
15,,00%	1 845 474	1 656 717	11,39%

Nous constatons que les méthodes ligne à ligne plus sont plus difficiles à mettre en place car elles nécessitent la construction de table d'expérience qui peuvent entrainer un biais dans l'estimation. De plus, elles sont moins stables lorsque le portefeuille manque d'information au niveau individuel tandis que les méthodes agrégées sont plus robustes et plus simples à mettre en place dans ce cas-là.

Conclusion

Les méthodes de provisionnement introduites dans ce mémoire ont eu pour but d'estimer les sinistres en cours de l'année N, la charge ultime et également de provisionner les tardifs qui ne sont pas connus de l'assureur au $31/12/N$ dans le cadre de l'arrêt de travail.

Les sinistres en cours ont été estimés à l'aide d'une approche ligne à ligne. Nous avons vu que cette méthode était précise grâce à une individualisation du risque. La méthode de Chain Ladder est moins précise et plus volatile à cause de la surparametrisation des coefficients de passages due à l'historique mais reste cependant correcte en ordre de grandeur.

En ce qui concerne l'estimation des tardifs, la méthode mis en place à l'aide de la loi d'incidence donne une estimation du nombre de tardifs assez correcte mais l'approche ligne à ligne utilisée ensuite sur ce nombre a eu pour effet de surestimer la charge à cause de l'hypothèse introduite.
Il aurait été intéressant de complexifier cette hypothèse en se basant sur une structure de l'ancienneté semblable à l'historique des sinistres clôs et de supposer que les tardifs ont la même représentation. Les résultats auraient peut être été plus précis et moins volatiles que la méthode de Chain Ladder.

Les deux méthodes ont leur avantage et leur inconvénient. La méthode ligne à ligne est certes plus précise mais elle demande une quantité d'informations au niveau individuel assez importante. Nous avons vu que lorsque cette information est manquante cela entraine un biais dans l'estimation. De plus, elle est plus difficile à mettre en place dans la pratique car elle nécessite la construction de tables d'expérience et cela peut également entrainer un biais dans l'estimation. La méthode de Chain Ladder quant à elle, est plus facile à mettre en place, ses résultats sont faciles à interpréter. Elle est assez robuste car elle n'est pas sensible à un manque d'information individuelle mais elle nécessite un historique important pour pouvoir être utilisée.

Synthesis note

Context and extent of the study

Insurance is a singular economic sector characterized by the inversion of its production cycle. Indeed, the insurer receives a premium from the insured and commits, in return, to pay him a claim in the event of an uncertain risk provided for in the contract during the coverage period.

Consequently, the main challenge for an insurance company is to estimate the amount of reserves that it will have to set aside in order to honor these commitments. Provisions are defined as the expected sum of discounted future cash flows added to the risk margin. They represent a highly significant portion of the insurance company's balance sheet liabilities. Therefore, they require a very precise assessment.

Throughout this research paper, we will be analyzing the results of various provisioning methods commonly used on the market regarding work stoppage. The framework of the study refers more specifically to temporary work disability (TWD), i.e. the transition of an individual from an active state to a state of incapacity by considering the duration of his stay in this state. The TWD's maximum duration 3 years, i.e. 36 months. Thus, the aim of the insurer is to estimate the residual duration of an individual in a state of disability in order to compensate him accordingly.

To do so, the insurer can use different provisioning methods. The methods generally used for work stoppage are line-by-line and aggregate provisioning methods based on runoff triangles. Each of these methods has pros and cons.

The purpose of this thesis is to analyze the results of these methods by comparing the estimated provisions with simulated data that best represent the work stoppage claim rate of a population of French working people. One method is considered more accurate than another if the difference between the provision and the simulation is smaller than the one obtained with the other method.

Process

The process set out in this paper follows the following plan:

- Construction of an age and gender structure that represents the French working population;
- From this structure, data simulation of the disability claims of each individual: date of occurrence, date of declaration, duration of maintenance in incapacity;
- Reserving, from the point of view of the insurer as of 31/12/N, the current and late claims using the different methods;
- Compare the difference between the estimated reserves and the actual duration of the simulation;
- Repeat the operation on 500 simulations to compare the expectation and empirical volatility of the reserves with the simulated durations.

We first construct an age and gender structure based on INSEE (French national institute for statistical and economic studies) data to accurately reflect the population of the French working population. The following age pyramid (in thousands) represents the shape of the structure.

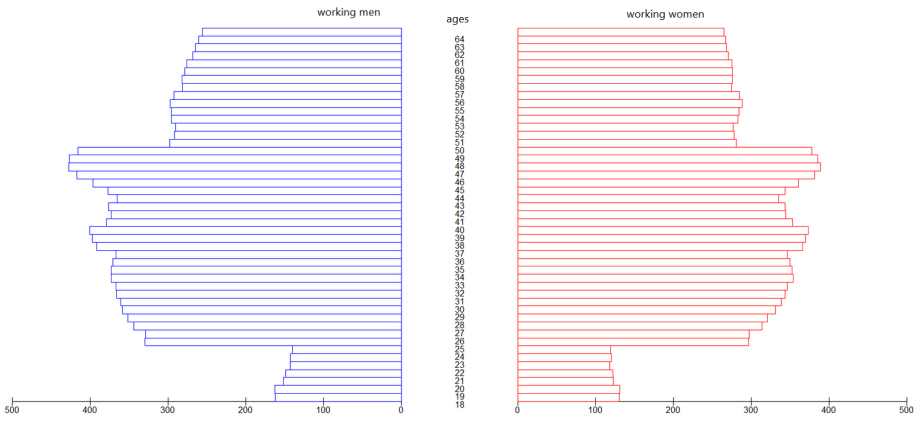


Figure 9: Age pyramid of the fictitious population of the French working population

From this structure, we simulate the work stoppage entries using the incidence law given by Prim'Act's experience table.
The table is constructed for loan insurance, which generally has a deductible of 3 months, hence it is important noting that the entries are considered with 3 months of seniority in disability.

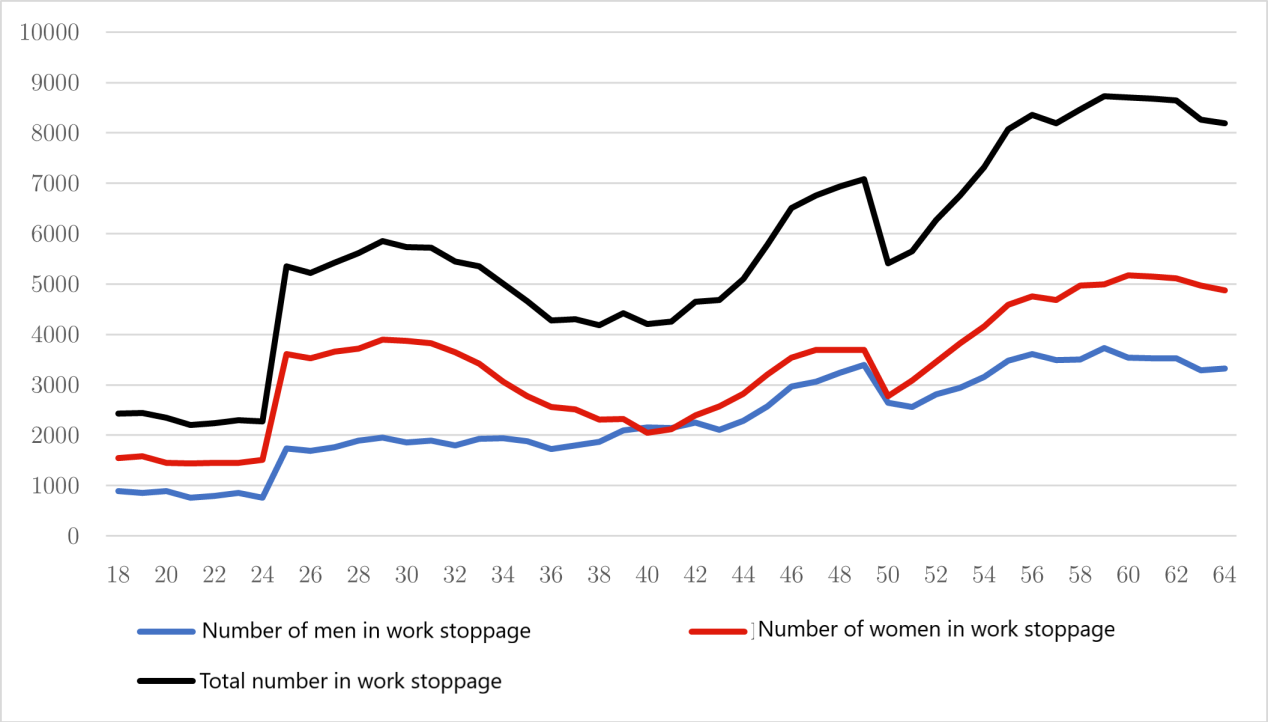
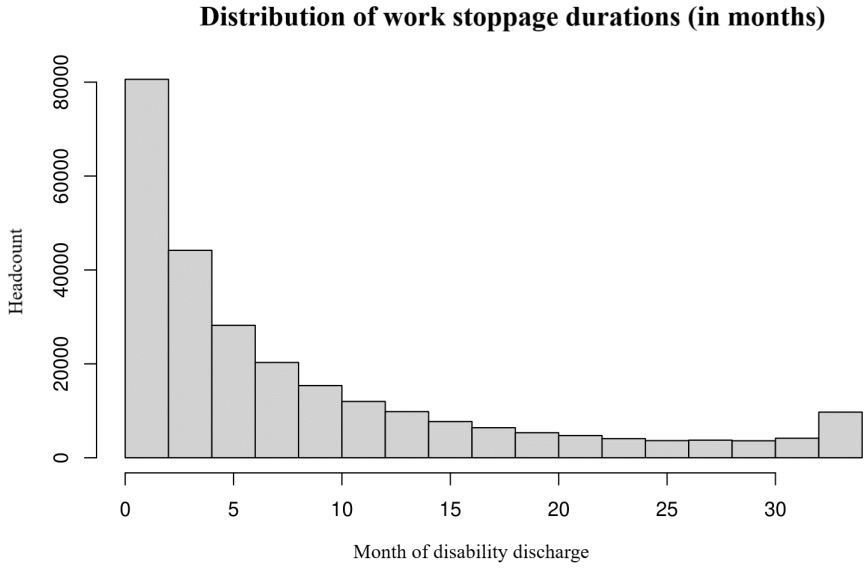


Figure 10: Total number of individuals (in thousands) in work stoppage (with 3 months of seniority)

We thus have a database with individuals who have been on a temporary work disability leave for at least 3 months. From this database, we simulate for each individual the characteristics of the claim, i.e. the date of occurrence, the date of declaration and the duration of the disability state.

The date of occurrence of claims is simulated over a year with a uniform distribution. The insurance declaration period is simulated with a log-normal distribution with a mean of 2 months. Lastly, the duration of the disability state is simulated using the maintenance law given by the BCAC table.

A duration of disability is simulated for each individual, according to his age. It is important to note that in the rest of the paper we will focus on the compensated duration. As a result, individuals have a duration of compensation that varies from 0 to 33 months because they have 3 months of seniority.



In this section, we place ourselves from the point of view of an insurer that seeks to provision for its work stoppage claims as of 31/12/N.

The insurer is unaware of the previous total durations. Therefore, his database was built from the simulation, whose stoppage durations are censored at 31/12/N and which only includes claims reported before 31/12/N according to the declaration period.

This database therefore contains 2 types of claims:

- claims whose compensation period ended before 31/12/N, which are called closed claims and for which the claim amounts are therefore known;
- claims for which the individual is still in a state of disability as of 31/12/N, which are called current claims and for which the claim amounts have yet to be estimated.

The aim of this insurer is therefore to set aside provisions for claims in progress as well as to estimate the cost of late claims of which the insurer is not yet aware as of 31/12/N. The cost of claims in the rest of this report is presented as a function of the number of compensated months.

For this purpose, we use 3 provisioning methods:

- a line-by-line provisioning method to estimate the cost of outstanding claims;
- an aggregated provisioning method: Chain Ladder based on a triangle in number of claims to estimate the ultimate cost of the year N and the late claims;
- a provisioning method based on the law of incidence to estimate late claims.

The line-by-line approach is based on the calculation of a provisioning coefficient in compensated months based on the seniority in the state and each individual's age.

The Chain Ladder method consists in the construction of a runoff triangle in number of claims thanks to a 3-year history generated in the same way as the previous simulation. The ultimate number of claims was then multiplied by an average cost of 7.62 reflecting the history to obtain an estimate of the cost.

The final provisioning method is based on incidence rates to estimate the number of late reported claims. Then we consider each late claim to have a seniority of 2 months of compensation because the delay in reporting claims is on average 2 months. We then perform a line-by-line approach on these claims.

Results on one simulation

The results obtained on a state of the world are as follows :

Indemnisation estimated (without IBNR)	Indemnisation simulated (without IBNR)	Difference in %
1 655 864	1 651 673	0,25%

Figure 11: Results of estimating current claims using the line-by-line method

Number of claims estimated	Number of claims simulated	Difference in %
270 051	262 758	2,78%
Ultimate charge estimated	**Ultimate charge estimated**	**Difference in %**
2 060 153	2 002 932	2,85%

Figure 12: Results of the estimation of the ultimate cost using the Chain Ladder method

The estimation of outstanding claims using the line-by-line method is fairly accurate with a deviation of 0.25%. This method remains accurate because the risk is individualized according to the characteristics of the individuals.

The Chain Ladder method gives a fairly accurate estimate of the ultimate cost with a deviation of 2.7%.

IBNR estimates

Chain Ladder method		
Number of IBNR estimated	Number of IBNR simulated	Difference in %
46 517	45 425	2,40%
Charge of IBNR estimated	Charge of IBNR simulated	Difference in %
402 174	346 215	16,16%
Method based on incidence law		
Number of IBNR estimated	Number of IBNR simulated	Difference in %
46 517	45 425	2,40%
Charge of IBNR estimated	Charge of IBNR simulated	Difference in %
427 764	346 215	23,55%

Figure 13: IBNR Estimation Results

The Chain Ladder method overestimates the number of IBNR with a deviation of 16.05%, while the method based on the law of incidence gives a fairly correct estimation with a deviation of 2.40%. Regarding the cost, the Chain Ladder method is however more accurate. Indeed, the rather strong hypothesis introduced on the seniority of the late reported claims in the 2nd method resulted in a rather important overestimation of the cost of IBNRs with a deviation of 23.55%.

Results on a sample of 500 simulations

The results obtained on 500 simulations are as follows :

	Indemnisation estimated (without IBNR)	Indemnisation simulated (without IBNR)
empirical mean	1 660 071,77	1 664 256,56
empirical deviation	5 054,41	5 648,72

Figure 14: Results of the line-by-line method over 500 simulations

	Total charge estimated	Total charge simulated
empirical mean	2 047 518,70	2 014 241,00
empirical deviation	25 716,05	6 269,44

Figure 15: Results of the Chain Ladder method over 500 simulations

The line-by-line method remains fairly accurate and sparsely volatile. The Chain Ladder method also remains fairly accurate for estimating the total cost but is much more volatile due to an irregularity in the history from year to year during the simulations.

IBNR estimates

	Charge of IBNR
empirical mean of simulation	349 984,43
charge of IBNR estimated with Chain Ladder method	
empirical mean	383 438,40
empirical deviation	24 982,65
difference with simulation	9,56%
charge of IBNR estimated with law of incidence method	
empirical deviation	421 542,37
empirical deviation	4 355,39
difference with simulation	20,45%

Figure 16: Methods results on the evaluation of IBNRs

Both methods tend to overestimate the cost of late reported claims. The Chain Ladder method is much more volatile (5 times more) than the other method. Indeed, since the other method is based on a line-by-line approach, it remains quite stable empirically. However, it is not very precise because of the introduced hypothesis.

We then compared the provisioning methods when the insurer makes an estimation error during the construction of its tables as well as when the insurer's portfolio presents a lack of individual information.
For this, shocks were introduced on tables used to estimate. We obtain the following results:

Impact of shoks made on incidence table

Shoks	Number of IBNR estimated	Number of IBNR simulated	Difference in %
10,00%	72 901	45 425	60,49%
5,00%	59 709	45 425	31,45%
2,00%	51 792	45 425	14,02%
1,00%	49 156	45 425	8,21%
0,50%	47 841	45 425	5,32%

Impact of shoks made on maintenance table

We notice that the shocks made on the incidence table lead to a large difference in the estimate of IBNR. Shocks made on maintain table lead to smaller deviations. The orders of magnitude of the shocks introduced are compared with what we fund in reality during the construction of tables. For this, we compared with various memories. The orders of magnitude seem consistent.

Finally, we have studied the impact of imperfect knowledge of the insurer's portfolio on the estimate of claims in progress. For this, we considered that the base of the insurer lacks information at an

Shoks	Charge estimated	Charge simulated	Difference in %
20,00%	1 817 672	1 656 717	9,72%
15,00%	1 776 112	1 656 717	7,21%
10,00%	1 734 553	1 656 717	4,70%
5,00%	1 692 994	1 656 717	2,19%

individual level. Database contains claims by age group, a number of each age group in workstoppage and an average seniority by age group.

In order to estimate charge, the insurer uses assumptions based on an average individual and then uses the line-to-line method on this one. Assumptions are based on the history of years $N-2$ and $N-1$. We use BCAC maintenance table and experience tables of the previous part to estimate the new base.
We obtain the following results :

Results of estimation with BCAC table

Charge estimée	Charge simulée	Ecart relatif
1 711 749	1 656 717	3,32%

Results of estimation with experience tables

Chocs	Charge estimée	Charge simulée	Ecart relatif
5,00%	1 756 324	1 656 717	6,01%
10,00%	1 800 899	1 656 717	8,70%
15,,00%	1 845 474	1 656 717	11,39%

We find that the line-to-lline methods are more difficult to set up than aggregate methods because they require the construction of experience tables which can lead to a bias in the estimation. Moreover, they are less stable when the portfolio lacks information at an individual level, while the aggregated methods are more robust and easier to set up in this case.

Conclusion

The purpose of the provisioning methods introduced in this paper was to estimate the current claims during the year N, the ultimate cost and also to provide for late reported claims that are not known to the insurer at 31/12/N in the context of work stoppage.

Current claims were estimated using a line-by-line approach. We have seen that this method was precise thanks to an individualization of the risk. The Chain Ladder method is less precise and more volatile because of the over-parameterization of the transition coefficients due to the history, but nevertheless remains correct in order of magnitude.

With regard to the estimation of the number of late reported claims, the method that was set up using the incidence law gives a fairly correct estimate of the number of late reported claims, but the line-by-line approach then used on this number resulted in an overestimation of the cost because of the hypothesis introduced.
It would have been interesting to increase the complexity of this hypothesis by using a seniority structure similar to the closed claims history and to assume that the late claims have the same representation. The results might have been more accurate and less volatile than the Chain Ladder method.

Both methods have their advantages and disadvantages. The line-by-line method is certainly more precise but it requires a rather large amount of information at the individual level. Moreover, it is more difficult to set up in practice because it requires the construction of experience tables and this can also lead to a bias in the estimation. Chain Ladder method is easier to set up, its results are easy to interpret. It is quite robust because it is not sensitive to a lack of individual information but it requires a significant history to be able to be used.

Remerciements

Je tiens tout d'abord à remercier mon tuteur en entreprise, Monsieur Killian PAYEN, consultant en actuariat chez Prim'Act, pour son soutien, ses conseils et le partage de son expérience professionnelle tout au long de mon stage. Je tiens également à remercier Maxime BEN-BRIK, consultant en actuariat chez Prim'Act, pour ses conseils et son soutien au cours de la rédaction de mon mémoire chez Prim'Act.

Par ailleurs, je souhaite exprimer toute ma gratitude à Monsieur Frédéric PLANCHET, associé chez Pim'Act, pour son expertise, sa disponibilité et son encadrement tout au long de la rédaction de ce mémoire.

Je remercie également Monsieur Geoffroy DELION de m'avoir permis de réaliser ce stage chez Prim'Act, ainsi que toute l'équipe pour leur accueil chaleureux et pour le partage de leurs expériences durant ce stage.

Mes remerciements s'adressent aussi à l'ensemble des enseignants et professionnels de l'ENSAE pour le partage de leurs connaissances dans le domaine de l'Actuariat, et en particulier, je remercie mon référent pédagogique Monsieur Christian-Yann ROBERT pour son suivi.

Enfin, je souhaite remercier ma famille pour son soutien tout au long de mes études et pour la relecture de mon mémoire.

Table des matières

Résumé **3**

Abstract **4**

Note de Synthèse **5**

Synthesis note **15**

Remerciements **25**

Table des matières **27**

Introduction **29**

1 DESCRIPTION DES DONNEES ET PERIMETRE DE L'ETUDE **31**

1.1 Cadre et périmètre de l'étude 31

1.2 Structure par âge de la population des actifs français 33

1.3 Description des tables d'incidence et de maintien utilisées 37

1.4 Lois de probabilité classiques utilisées pour la simulation 45

2 Simulation de la sinistralité d'arrêt de travail **47**

2.1 Méthodologie de la simulation 47

2.2 Simulation de l'incidence en arrêt de travail 50

2.3 Simulation des caractéristiques du sinistre d'arrêt de travail 52

3 Méthodes de provisionnement **57**

3.1 Les provisions en assurance non vie 57

3.2 Méthode de provisionnement ligne à ligne 60

3.3 Méthode de provisionnement agrégée : la méthode de Chain Ladder 63
3.4 Méthode de provisionnement des IBNR à l'aide de la loi d'incidence 69
3.5 Résultats des différentes méthodes de provisionnement sur 500 simulations 72

4 Etude de la stabilité des méthodes de provisionnement suite à une erreur de modélisation et une connaissance imparfaite du portefeuille assuré 81

4.1 Impact de l'erreur d'estimation des tables sur le provisionnement 81
4.2 Impact d'une structure imparfaite du portefeuille sur le provisionnement 87
4.3 Conclusion sur la stabilité des méthodes 88

Conclusion 89

Bibliographie 91

Introduction

L'assurance est un secteur économique singulier qui est caractérisé par une inversion de son cycle de production. En effet, l'assureur reçoit de la part de son assuré une prime et s'engage, en contrepartie, à lui verser une prestation en cas de réalisation d'un risque aléatoire prévu dans le contrat pendant la période de couverture. L'assureur doit donc constituer des provisions afin d'honorer ses engagements.

En 2019, le montant des provisions techniques des contrats d'assurance-vie s'élève à 1788 milliards d'euros et 169 milliards d'euros pour les sociétés d'assurance non vie d'après un rapport de la Fédération Française de l'Assurance. Les provisions représentent une part très importante du passif au bilan de la compagnie d'assurance. Elles nécessitent donc une évaluation très précise.

Ainsi, le principal enjeu pour une compagnie d'assurance est d'estimer le montant des réserves qu'elle aura à constituer afin d'honorer ses engagements en cas de réalisation du sinistre prévu par la police pendant la période de couverture. La difficulté du provisionnement réside dans la prédiction de la prestation à verser qui présente un caractère incertain. L'assureur a donc recours à différentes méthodes de provisionnement afin d'évaluer cette charge.

L'objectif de ce mémoire est d'analyser la performance des différentes méthodes de provisionnement couramment utilisées sur le marché dans le cadre de l'arrêt de travail. Le cadre de l'étude se concentre plus particulièrement sur l'incapacité temporaire de travail (ITT). L'incapacité temporaire de travail (ou ITT) désigne la période pendant laquelle un individu actif victime d'un accident ou d'une maladie ne peut pas occuper son emploi. Le rôle de l'assureur est donc de couvrir cette perte et donc d'estimer la durée résiduelle d'un individu en arrêt de travail afin de l'indemniser.

Les méthodes traditionnellement utilisées dans le cadre de l'arrêt de travail sont généralement les méthodes de provisionnement ligne à ligne qui nécessitent l'utilisation de tables d'expériences ou réglementaires et les méthodes de provisionnement agrégées à base de triangles de règlement.

Afin d'analyser la performance de ces différentes méthodes, nous construisons une structure par âge et par sexe qui représente la population des actifs français. A partir de cette structure, nous simulons des entrées en arrêt de travail avec leur durée totale. Nous nous retrouvons donc avec une réalité simulée d'individus en arrêt de travail dont la durée totale d'incapacité est connue.

Nous nous plaçons ensuite du point de vue de l'assureur qui cherche à provisionner ses sinistres d'arrêt de travail au $31/12/N$ et qui utilise donc les méthodes de provisionnement afin d'estimer les durées résiduelles. Les charges estimées par les méthodes de provisionnement sont ensuite confrontées avec la

simulation.

Dans la première partie de ce mémoire, nous décrivons les données et le périmètre de l'étude. Dans cette perspective, tous les éléments qui serviront à la simulation ainsi que le contexte seront détaillés.

Dans la seconde partie, nous expliquons les étapes de la simulation réalisée à partir des données décrites dans la première partie. La cohérence des résultats de cette simulation sera également analysée.

La troisième partie concerne le point de vue de l'assureur, à savoir l'utilisation des méthodes de provisionnement et la confrontation des résultats avec la simulation. A la fin de cette partie, le processus est réalisé sur un échantillon de 500 simulations afin d'obtenir une distribution de la provision des différentes méthodes.

Enfin la dernière partie concerne l'étude de la stabilité des méthodes de provisionnement lorsque l'assureur construit ses propres tables d'expérience et lorsque celui-ci possède un portefeuille avec des données incomplètes.

Chapitre 1

DESCRIPTION DES DONNEES ET PERIMETRE DE L'ETUDE

1.1 Cadre et périmètre de l'étude

1.1.1 La Prévoyance en France

La prévoyance désigne l'assurance qui permet de couvrir les risques liés à la personne. Selon l'Art. de loi $n°89 - 1009$ du 31 décembre 1989, dite loi EVIN : "La prévoyance regroupe les opérations ayant pour objet la prévention et la couverture du risque décès, des risques portant atteinte à l'intégrité physique de la personne ou liés à la maternité ou des risques d'incapacité de travail ou d'invalidité ou du risque de chômage".

Le système de prévoyance en France se décompose en plusieurs catégories :

Le régime de base de la Sécurité Sociale

Le régime de base de la Sécurité Sociale permet de couvrir les risques liés aux problèmes de santé d'un individu. Cette indemnisation reste cependant partielle.

Le régime complémentaire obligatoire

Une prévoyance complémentaire n'est généralement pas obligatoire mais elle le devient lorsqu'une convention collective la prévoit.

La prévoyance collective

La prévoyance collective désigne un contrat de prévoyance souscrit par l'entreprise pour ses salariés. Cela peut avoir plusieurs avantages tels que de meilleurs tarifs grâce à la mutualisation des risques sur l'ensemble des salariés de l'entreprise. De plus, une partie des cotisations peut être payée par l'entreprise et le contrat est adapté au métier des salariés.

Ce contrat peut être obligatoire ou facultatif en fonction du secteur.

La prévoyance individuelle

La prévoyance individuelle est totalement facultative, c'est une prévoyance complémentaire qui permet de compléter le régime de base de la Sécurité Sociale.

Les garanties de la Prévoyance

Les risques couverts par la Prévoyance se décomposent en 3 catégories : la garantie décès et les garanties incapacité et invalidité de travail.

La garantie décès se caractérise par le versement d'un capital ou d'une rente aux bénéficiaires du contrat en cas de décès. Les frais d'obsèques peuvent être également indemnisés.

Les garanties d'incapacité et d'invalidité constituent la garantie d'arrêt de travail que nous détaillons dans la partie suivante.

1.1.2 L'arrêt de travail

L'arrêt de travail désigne l'impossibilité pour un individu d'exercer son activité professionnelle. Ce risque entraine une perte de revenus pour l'actif et l'assureur propose donc de couvrir cette perte.

L'arrêt de travail se scinde en deux catégories : l'incapacité temporaire de travail (ITT) et l'invalidité (ou IPT : incapacité permanente de travail). L'incapacité temporaire est la période pendant laquelle l'assuré victime d'un accident ou d'une maladie est dans l'incapacité d'occuper son emploi.

Cela signifie que l'assuré ne peut plus travailler pendant un certain temps. On parle alors d'ITT (incapacité temporaire de travail). L'incapacité a une durée maximale de 36 mois (soit 3 ans). Si le délai est supérieur à ce seuil, on parle alors d'invalidité.

L'invalidité (ou IPT) est un état physique ou mental mettant un individu dans l'impossibilité totale d'exercer son activité professionnelle. Etant un état définitif, la garantie d'invalidité reste cependant valable seulement avant l'âge de la retraite.

Ce mémoire se concentre uniquement sur l'incapacité temporaire de travail, à savoir l'entrée en état d'incapacité d'un individu, ainsi que sur sa durée d'arrêt dans cet état (maximum 36 mois).

Cas particulier : l'assurance emprunteur

L'assurance emprunteur est une assurance qui couvre la totalité ou une partie du remboursement d'un crédit en cas de décès, incapacité, invalidité et perte d'emploi.

Il est important de noter pour la suite du mémoire que les franchises des contrats d'assurance emprunteur d'arrêt de travail sont généralement de 3 mois.

L'objectif de ce mémoire est de simuler une population fictive en arrêt de travail à partir d'une structure par âge de la population des actifs français, de construire une base relevant de cette simulation afin de tester les différentes méthodes de provisionnement usuelles et de pouvoir les comparer à la simulation.

1.2 Structure par âge de la population des actifs français

Afin de simuler les données, il est nécessaire de partir d'une structure par âge qui représente au mieux une population d'assurés français. Puisqu'il s'agit d'une modélisation de l'arrêt de travail, nous considèrons une base de données de l'INSEE qui représente la population d'actifs français en 2017 par tranche d'âge et par sexe.

Âge	Femmes	Hommes	Total
15 à 24 ans	1248	1532	2780
25 à 49 ans	8703	9435	18138
50 à 64 ans	4142	4228	8370

Source : *INSEE, enquête emploi*

FIGURE 1.1 : Population active par sexe et âge en 2017 (en milliers)

Puisque cette structure est sous forme de tranche d'âge (de 18 à 64 ans), nous reconstituons une population d'actifs français fictive à partir de la base de donnée de la population totale française de l'INSEE de 2020 (voir ci-dessous).

Année de naissance	Âge révolu	Nombre d'hommes	Nombre de femmes	Ensemble
2019	0	343009	329909	672918
2018	1	348625	333151	681776
2017	2	354653	340515	695168
2016	3	366334	350669	717003
2015	4	376914	360433	737347
2014	5	386296	374278	760574
2013	6	392050	375104	767154
2012	7	401732	381332	783064
2011	8	402835	386329	789164
2010	9	412838	395927	808765
2009	10	410032	390752	800784
2008	11	411597	392913	804510
2007	12	406858	389218	796076
2006	13	414061	396573	810634
2005	14	405934	386107	792041
2004	15	405931	386046	791977
2003	16	405172	385143	790315
2002	17	405928	383887	789815

Source : *INSEE, enquête emploi*

FIGURE 1.2 : Extrait de la population totale par sexe et âge au 1er janvier 2020, France métropolitaine

Ainsi, nous reconstruisons une base d'actifs français âge par âge à l'aide des deux bases précédentes de manière proportionnelle.

Âge	Nombre d'hommes	Nombre de femmes	Ensemble
18	162047	130003	292050
19	162862	130412	293274
20	151875	122710	274585
21	149216	121965	271181
22	142910	117919	260829
23	142719	119567	262286
24	140093	118667	258761
25	329281	295915	625196
26	328589	296872	625461
27	344231	313331	657563
28	351712	320347	672058
29	358735	330504	689239
30	360401	338242	698642
31	366500	343411	709911
32	366901	345972	712872
33	373089	353614	726703
34	372806	352081	724888
35	370849	349534	720382
36	366786	345725	712511
37	391941	366113	758054
38	396988	369305	766293
39	400490	372691	773181
40	379061	352893	731954
41	373213	343767	716980
42	376354	342900	719255
43	365337	334652	699989
44	376906	342802	719709
45	396633	360633	757266
46	417522	380801	798323
47	427630	388459	816089
48	426884	385291	812175
49	415962	377343	793305
50	297553	280420	577974
51	291826	277808	569634
52	290258	276258	566516
53	295701	282774	578475
54	295745	284065	579809
55	297339	287847	585186
56	292233	284338	576571
57	281624	274478	556102
58	281994	275531	557525
59	278635	275239	553874
60	275543	274684	550226
61	268125	269891	538015
62	264779	267955	532734
63	260350	266258	526609
64	256095	264754	520849

FIGURE 1.3 : Population fictive d'actifs français par sexe et âge

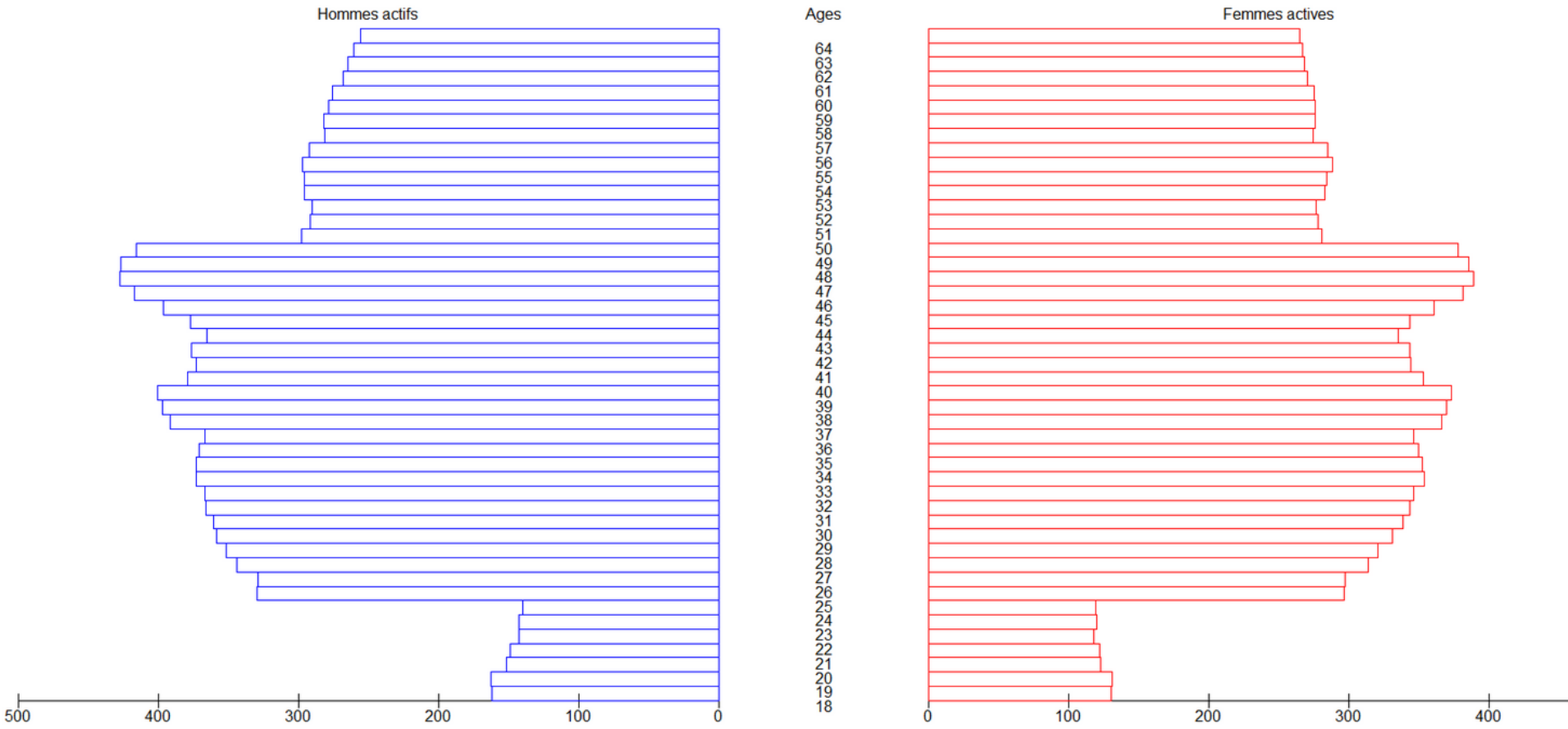


FIGURE 1.4 : Pyramide des âges de la population fictive des actifs français

Nous observons que le nombre d'hommes et de femmes en activité entre 18 et 24 ans est le plus faible, environ en moyenne 150 000 pour les hommes et 110 000 en ce qui concerne les femmes.

A partir de 25 ans, le nombre d'hommes et de femmes en activité augmente fortement et est environ 2,5 fois plus élevé que la tranche d'âge précédente (environ 300 000) de chaque côté.

Entre 25 et 49 ans, le nombre d'hommes et de femmes en activité correspond environ à une moyenne de 350 000 ; avec un maximum de 420 000 d' hommes et 380 000 de femmes actives à l'âge de 47 ans.

A partir de 50 ans, jusqu'à l'âge de 64 ans, le nombre d'hommes et de femmes en activité diminue pratiquement d'un quart et reste en moyenne à 300 000 du coté des hommes et 280 000 du côté des femmes.

Le total de personnes en activité de cette population fictive est d'environ 28 millions, ce qui semble cohérent puisque le nombre de personnes actives en 2017 est d'environ 29 millions.
Cet écart peut s'expliquer par le fait que nous ne prenons pas en compte les âges de 15 à 17 ans. Cette population semble donc correcte et constitue la base principale pour modéliser l'arrêt de travail à l'aide des différentes tables qui sont décrites dans la partie suivante.

1.3 Description des tables d'incidence et de maintien utilisées

Afin de modéliser et de simuler la survenance des sinistres d'arrêt de travail ainsi que leur durée sur la structure précédente, nous utilisons deux tables :

- La table d'incidence de Prim'act qui est une table d'expérience ;
- La table réglementaire du BCAC (Bureau Commun des Assurances Collectives)

Dans cette partie, nous décrivons ces deux tables après avoir rappelé le cadre de l'étude qui concerne les modèles de durée.

1.3.1 Cadre de l'étude : Les modèles de durée

Lorsque l'on s'intéresse à des variables de durée, on utilise un cadre de modélisation spécifique. En effet, le cadre des modèles de durée est utilisé dans de nombreux domaines de l'assurance : durée de vie humaine, durée d'arrêt de travail, durée de chômage etc ...

Les durées ne sont pas des variables aléatoires ordinaires, elle contiennent des spécificités. C'est pourquoi on a besoin d'une certaine modélisation.
En effet, une durée est une variable positive dans $[0, +\infty[$. Bien qu'on puisse utiliser le logarithme de cette variable pour se débarrasser de cette contrainte, les distributions utilisées pour décrire le phénomène de durée sont différentes des distributions classiquement utilisées (ex : loi normale).

De plus, les durées sont difficilement observables dans leur intégralité, et les observations incomplètes peuvent introduire un biais statistique. Enfin, l'interprétation de la variable aléatoire de durée se fait à l'aide de la fonction de survie ou de hasard et non pas avec la fonction de répartition comme c'est le cas pour d'autres variables.

On note T la variable aléatoire de durée, à valeurs dans $[0, +\infty[$.

La densité de T est définie comme $\forall t \in \mathbb{R}_+, f(t) = \lim_{dt \to 0} \dfrac{P(T \in [t, t+dt])}{dt}$

La fonction de répartition F de T est définie comme $\forall t \in \mathbb{R}_+, F(t) = \mathbb{P}(T \leqslant t)$

Fonction de survie

Définition : La fonction de survie S d'une variable aléatoire T est :

$$\begin{aligned} \forall t \in \mathbb{R}_+, \quad S(t) &= \mathbb{P}(T \geqslant t) \\ &= 1 - F(t-) \end{aligned} \tag{1.1}$$

Fonction de survie conditionnelle

Définition : La fonction de survie conditionnelle d'une variable aléatoire T sachant T$\geqslant x$ est :

$$\begin{aligned} S(t \mid x) &= P(T > u + t \mid T > u) \\ \text{pour t} \geqslant x, \quad S(t \mid x) &= \frac{S(t)}{S(x)} \end{aligned} \tag{1.2}$$

Espérance de vie résiduelle

Définition : L'espérance de vie résiduelle à l'âge x est :

$$\begin{aligned} e(x) &= E(T - x \mid T \geqslant x) \\ &= \int_0^\infty S(t \mid x)dt - x \\ &= \frac{1}{S(x)} \int_x^\infty S(t)dt \end{aligned} \tag{1.3}$$

Taux de risque instantané (ou fonction de hasard)

Définition :

- **Cas discret :** T prend ses valeurs dans un ensemble dénombrable, $\{t_1, \ldots, t_n, \}$. Le taux de risque instantané μ est défini par :

$$\begin{aligned} \forall t, \quad \mu(t) &= \mathbb{P}(T = t \mid T \geqslant t) \\ &= \frac{\mathbb{P}(T = t)}{\mathbb{P}(T \geqslant t)} \end{aligned} \tag{1.4}$$

  Nous en déduisons une relation importante (qui sera utilisée en pratique plus tard) :

$$S(t) = \prod_{i|t_i<t} \left(1 - \mu\left(t_i\right)\right) \tag{1.5}$$

  Ce qui signifie qu'à partir du taux de risque instané, nous retrouvons la fonction de survie S et donc la loi de T.

- **Cas continu :** Le taux de risque instantané μ est défini par :

$$\begin{aligned} \mu(t) &= \lim_{dt \to 0^+} \frac{\mathbb{P}(T \in [t, t + dt] \mid T \geq t)}{dt} \\ &= \frac{f(t)}{S(t)} \end{aligned} \tag{1.6}$$

  Nous en déduisons la relation :

$$S(t) = \exp\left(-\int_{-\infty}^{t} \mu(u)du\right) \tag{1.7}$$

Le taux de risque instantané jouera le rôle des q_x (en notation actuarielle) des tables actuarielles que nous définissons dans la partie suivante.

Notations et tables actuarielles

Dans ce qui suit, nous prenons l'exemple d'une table de mortalité pour introduire les notations de façon simple . Ces notations serviront ensuite pour décrire notre cadre d'étude sur l'arrêt de travail dans la partie suivante.
Une table de mortalité est présentée de la façon suivante :

Age x	Lx
0	100000
1	99616
2	99583
3	99562
4	99545
5	99531
6	99519
7	99508
8	99498
9	99488
10	99478
11	99467
12	99456
13	99444
14	99431
15	99415
16	99395
17	99371
18	99342
19	99309
20	99274
21	99239
22	99205
23	99171
24	99137
25	99103
26	99068
27	99033
28	98997
⋮	⋮

FIGURE 1.5 : Extrait de la table TH-00-02 décès

On note L_x le nombre de survivants moyen à l'âge x. Nous partons d'un nombre de départ de 100 000 survivants à l'âge 0. Puis chaque année, le nombre de survivants diminue.
La fonction de survie se définit par la relation suivante :

$$S(x) = \frac{L_x}{L_0} \tag{1.8}$$

Nous notons q_x la probabilité que la durée de vie T soit comprise entre x et $x+1$ sachant que $T \geq x$:

$$\begin{aligned} q_x &= \mathbb{P}(x \leqslant T < x+1 \mid T \geqslant x) \\ &= \frac{\mathbb{P}(\{x \leqslant T < x+1\} \cap \{T \geqslant x\})}{\mathbb{P}(T \geqslant x)} \\ &= \frac{S(x) - S(x+1)}{S(x)} \\ &= 1 - \frac{S(x+1)}{S(x)} \end{aligned} \tag{1.9}$$

Nous avons donc, d'après la relation précédente :

$$\begin{aligned} q_x &= \frac{L_x - L_{x+1}}{L_x} \\ &= 1 - \frac{L_{x+1}}{L_x} \end{aligned} \tag{1.10}$$

L'arrêt de travail dans le cadre des modèles de durée

En ce qui concerne l'arrêt de travail, nous considérons le modèle d'états suivant :

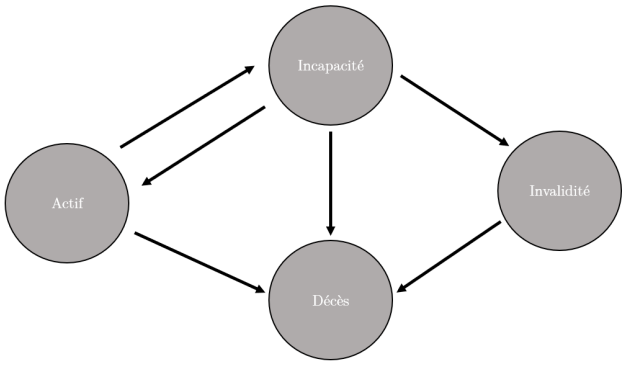


FIGURE 1.6 : Modélisation des états dans le cadre de l'arrêt de travail

Dans le cadre de ce mémoire, nous nous concentrons sur l'état incapacité (le passage d'actif à décès et invalidité à décès ne nous intéresse pas dans le cadre de ce mémoire). Deux aspects sont donc à prendre en compte :

- Un individu d'âge x dans l'état actif a une probabilité non nulle de rentrer dans l'état d'incapacité. Cette probabilité est la probabilité d'incidence notée q_x^{inc} ;

- Un individu d'âge x dans l'état d'incapacité a une probabilité non nulle de sortir de cet état (3 possibilités : retourner dans l'état actif, passer dans l'état invalide, passer dans l'état décès).

Nous nous intéressons donc à la probabilité de maintien dans cet état, notée q_x^{main}

Ces probabilités sont données par des tables actuarielles. La loi d'incidence sera donnée par la table d'expérience de Prim'Act et la loi de maintien par la table réglementaire du BCAC.

1.3.2 Table d'incidence Prim'act

Tables d'expérience

Les tables réglementaires reflètent l'ensemble de la population française, or cette population présente une hétérogénéité et donc représente mal le comportement d'une population spécifique d'un portefeuille d'assurance.

C'est pourquoi, il est intéressant pour les assurances de construire leur propre table d'expérience pour refléter au mieux la population assurée qu'elles couvrent.

Dans le cadre de ce mémoire, nous utilisons la table d'incidence Prim'Act afin de simuler les entrées en arrêt de travail ainsi que l'estimation des provisions pour les sinistres tardifs.

Tables d'incidence Prim'act

La loi d'incidence est donnée grâce à la table d'incidence Prim'Act. Cette table donne une estimation des taux d'entrée en incapacité pour des contrats de type assurance emprunteur.
Dans le cas de ce type de contrat, la franchise est souvent de 90 jours. C'est à dire que la garantie est acquise mais l'indemnité s'effectue le 91ème jour.

Les taux d'entrée à l'âge x q_x^{inc} sont donc construits en prenant en compte ce délai de franchise. Autrement dit, les taux d'entrée correspondent à la probabilité, pour un individu, de rentrer en incapacité le 4ème mois sachant qu'il était en état d'incapacité pendant 3 mois.

Cette information est importante à prendre en compte pour la simulation de la durée de maintien ainsi que pour le provisionnement.
La table d'incidence de Prim'Act se présente sous cette forme :

Age	Homme	Femme
18	$0,5364\%$	$1,2129\%$
19	$0,5364\%$	$1,2129\%$
20	$0,5364\%$	$1,2129\%$
21	$0,5364\%$	$1,2129\%$
22	$0,5364\%$	$1,2129\%$
23	$0,5364\%$	$1,2129\%$
24	$0,5364\%$	$1,2129\%$
25	$0,5364\%$	$1,2129\%$
26	$0,5315\%$	$1,1987\%$
⋮	⋮	⋮

FIGURE 1.7 : Extrait de la table d'incidence de Prim'act

Le graphique ci-dessous permet une visualisation plus claire de l'évolution des taux en fonction de l'âge et du sexe.

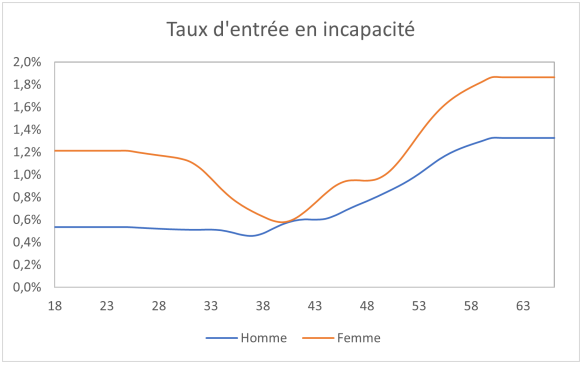


FIGURE 1.8 : Loi d'incidence Prim'Act

Nous observons que les taux d'entrée en incapacité sont plus importants chez les femmes que chez les hommes. Quasiment le double pour les âges de 18 à 33 ans.

Ils diminuent jusqu'à 0.6% à l'âge de 39 ans pour atteindre le même taux que celui des hommes puis augmentent jusqu'à $1,9\%$ à l'âge de 60 ans.

Au même titre que les femmes, les taux des hommes restent constants entre 18 et 33 ans, mais sont cependant deux fois moins importants : autour de $0,6\%$. Puis, ils augmentent jusqu'à $1,3\%$ à l'âge de 60 ans.

Cette table servira donc à simuler la loi d'incidence en arrêt de travail de notre population active. Nous verrons les étapes de cette simulation dans le chapitre suivant.

Le deuxième aspect à prendre en compte dans le cadre de l'étude, est la probabilité de maintien d'un individu d'âge x dans l'état d'incapacité. Cette loi est donnée grâce à la table réglementaire du BCAC.

1.3.3 Table de maintien réglementaire : BCAC 2013

La table BCAC de maintien en incapacité se présente sous cette forme :

	0	**1**	...	**35**	**36**
20	10000	4414		24	18
21	10000	4638		23	17
22	10000	4848		22	17
23	10000	5026		32	25
24	10000	5159		40	33
25	10000	5273		41	33
26	10000	5357		41	33
27	10000	5427		46	39
28	10000	5477		50	42
29	10000	5526		54	46
30	10000	5567		54	46
...			...		...
55	10000	6690		171	146
56	10000	6778		182	156
57	10000	6845		204	176
58	10000	6872		232	203
59	10000	6838		259	230
60	10000	6777		277	247
61	10000	6747		283	255
62	10000	6710		264	235
63	10000	6698		226	197
64	10000	6668		198	171
65	10000	6647	...	233	213

FIGURE 1.9 : Extrait de la table de maintien en incapacité BCAC 2013

Il s'agit d'une table à 2 dimensions : les lignes correspondent aux âges auxquels un individu entre dans l'état d'incapacité, les colonnes correspondent aux mois d'ancienneté de l'individu passé dans l'état d'incapacité.

Dans cette table, les âges vont de 20 à 65 ans et la durée maximale d'incapacité étant de 3 ans, les mois d'ancienneté vont de 0 à 35.

Ainsi, $L_{x,t}$ correspond au nombre d'individus d'âge x, dont l'ancienneté est t mois. Cet effectif commence à 10 000 et décroit au fur et à mesure que l'ancienneté augmente.

À partir de cette loi de maintien, nous en déduisons donc les taux $q_{x,t}^{main}$ qui correspondent à la probabilité de sortie d'un individu rentré en incapacité à l'âge x entre le t et le t+1 ème mois.
Nous notons T_x , la durée de maintien en incapacité d'un individu dont la survenance a eu lieu à l'âge x.

$$\begin{aligned} q_{x,t}^{main} &= \mathbb{P}(t \leqslant T_x < t+1 \mid T_x \geqslant t) \\ &= \frac{L_{x,t} - L_{x,t+1}}{L_{x,t}} \end{aligned} \tag{1.11}$$

Le graphique des taux de sortie a l'allure suivante :

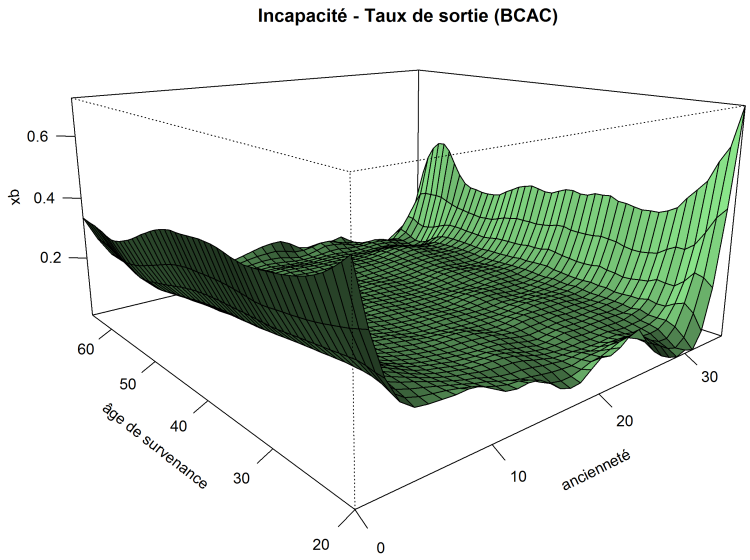


FIGURE 1.10 : Taux de sortie en incapacité BCAC 2013

Ces taux nous permettront de simuler les durées complètes d'arrêt de travail de la population ainsi que de provisionner les sinistres.

Il est important de noter que puisque la table d'incidence de Prim'Act fournit des taux d'entrée en incapacité avec 3 mois d'ancienneté, les 3 premières colonnes de la table du BCAC seront retirées pour simuler les durées d'arrêt indemnisées ainsi que pour provisionner.

Dans le cadre de ce mémoire, ce sont ces deux tables que nous utilisons pour modéliser l'arrêt de travail. Mais nous nous servons également de lois plus usuelles en ce qui concerne la simulation.

1.4 Lois de probabilité classiques utilisées pour la simulation

1.4.1 Loi uniforme

Soient deux réels a et b tels que $a < b$.
X suit une loi uniforme sur l'intervalle $[a; b]$ signifie que la densité de probabilité de X est la fonction f définie par :

$$f(x) = \begin{cases} 0 & \text{si } x < a \\ \dfrac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{si } x > b \end{cases} \tag{1.12}$$

1.4.2 Loi log-normale

Rappel : loi normale

La loi normale est l'une des lois de probabilité les plus utilisées en pratique.
X suit une loi normale de paramètres μ (espérance) et σ (écart-type), notée $\mathcal{N}\left(\mu, \sigma^2\right)$ si sa densité de probabilité est définie sur $\mathbb{R}$ par la fonction f :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{1.13}$$

Loi log-normale

X suit une loi log-normale de paramètre μ et σ si la variable $Y = ln(X)$ suit une loi normale d'espérance μ et d'écart-type σ.
Sa densité de probabilité est définie sur $]0; +\infty[$ par la fonction f :

$$f_Y(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) = \frac{1}{x} f_X(\ln(x)) \tag{1.14}$$

Ces lois de probabilité serviront à simuler les caractéristiques intrinsèques d'un sinistre d'assurance que nous détaillerons dans le chapitre suivant.

Conclusion

Ce chapitre était un inventaire des outils utilisés dans le cadre de ce mémoire afin de simuler les sinistres d'arrêt de travail de la population fictive d'actifs que nous avons créé. Nous allons voir au cours du chapitre suivant comment nous nous servons de ces outils dans le cadre de la simulation.

Chapitre 2

Simulation de la sinistralité d'arrêt de travail

2.1 Méthodologie de la simulation

Notre objectif est d'effectuer une simulation des arrêts de travail de la population fictive des actifs français afin de pouvoir réaliser les différentes techniques de provisionnement sur cette simulation. L'estimation des méthodes sera ensuite confrontée à la charge des sinistres de la simulation.

L'organigramme de la page suivante décrit le processus mis en place de façon synthétique.

La méthodologie employée comprend 3 étapes qui seront détaillées de façon plus précise dans la suite du mémoire.

Etape 1 : Simulation d'une population fictive en arrêt de travail

À partir de la structure par âge et par sexe des actifs français construite précédemment, nous simulons aléatoirement des entrées en arrêt de travail avec la méthode d'inversion (que nous détaillons dans la section suivante) à l'aide des taux fournis par la table d'incidence de Prim'Act.

Nous obtenons ainsi une base qui contient un nombre d'individus en arrêt de travail âge par âge (de 18 à 64 ans) avec une ancienneté de 3 mois.
Pour chaque individu, nous générons aléatoirement les caractéristiques du sinistre d'assurance : la date de survenance, la date de déclaration ainsi que la durée totale de maintien en incapacité.

Nous nous retrouvons donc avec une base sur une année N comportant des sinistres d'arrêt de travail dont leur durée totale d'incapacité est connue.
Nous simulons de la même manière deux autres bases sur les années $N-1$ et $N-2$ pour constituer un historique qui servira pour le provisionnement.

Etape 2 : Construction d'une base d'assurance à partir de la simulation

Dans cette étape, nous nous plaçons du point de vue de l'assureur au $31/12/N$. Celui-ci ne connait que les sinistres dont la date de déclaration est antérieure au $31/12/N$ et cherche à estimer les tardifs. De plus à cette date, certains sinistres sont en cours et d'autres sont clos (nous détaillons cette étape dans le chapitre 3). La base de l'assureur est donc construite à partir de la simulation afin de respecter ces critères. Les sinistres de la base sont connus de l'assureur et les durées d'arrêt des sinistres en cours

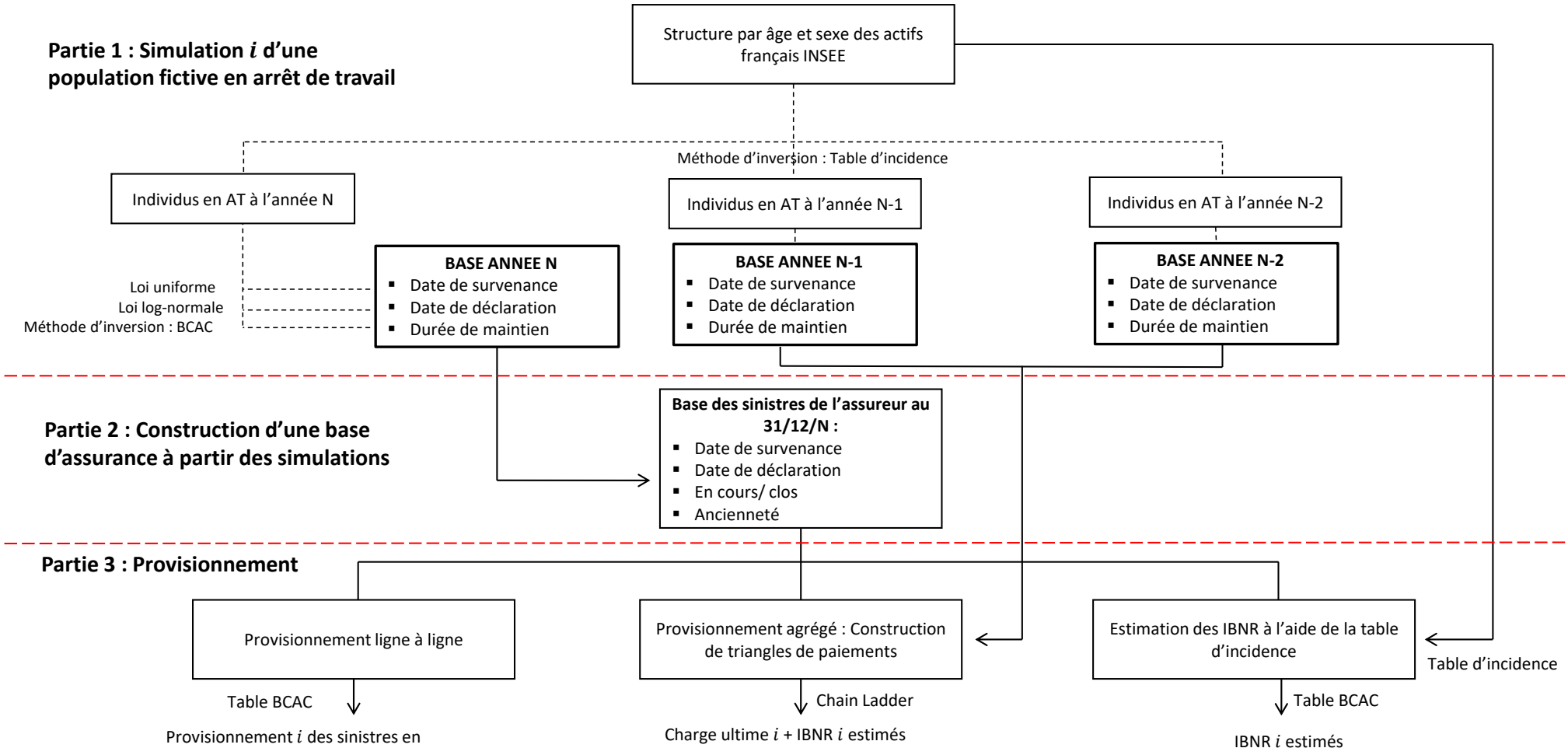
sont censurées au $31/12/N$. Le but est maintenant de provisionner cette base.

Etape 3 : Provisionnement de la base de l'assureur

L'assureur utilise 3 méthodes afin de provisionner sa base : la méthode de provisionnement ligne à ligne pour estimer la charge des sinistres en cours, une méthode de provisionnement agrégée à base de triangles de paiements (Chain Ladder) pour estimer la charge ultime ainsi que les tardifs et enfin une méthode basée sur la loi d'incidence pour estimer les tardifs.

À la fin, l'objectif est de répéter le processus décrit 500 fois afin d'obtenir une distribution empirique des estimations des méthodes.

Nous représentons les étapes au sein de l'organigramme suivant :


Partie 1 : Simulation i d'une population fictive en arrêt de travail
Structure par âge et sexe des actifs français INSEE
Méthode d'inversion : Table d'incidence
Individus en AT à l'année N
Individus en AT à l'année N-1
Individus en AT à l'année N-2
Loi uniforme
Loi log-normale
Méthode d'inversion : BCAC
BASE ANNEE N
Date de survenance
Date de déclaration
Durée de maintien
BASE ANNEE N-1
Date de survenance
Date de déclaration
Durée de maintien
BASE ANNEE N-2
Date de survenance
Date de déclaration
Durée de maintien
Partie 2 : Construction d'une base d'assurance à partir des simulations
Base des sinistres de l'assureur au 31/12/N :
Date de survenance
Date de déclaration
En cours/ clos
Ancienneté
Partie 3 : Provisionnement
Provisionnement ligne à ligne
Provisionnement agrégé : Construction de triangles de paiements
Estimation des IBNR à l'aide de la table d'incidence
Table d'incidence
Table BCAC
Provisionnement i des sinistres en cours
Chain Ladder
Charge ultime i + IBNR i estimés
Table BCAC
IBNR i estimés

A partir de la structure par âge et sexe de la population que nous avons décrite dans la partie précédente, la première étape de simulation concerne la simulation des entrées en incapacité.

2.2 Simulation de l'incidence en arrêt de travail

2.2.1 Méthode d'inversion

Afin de générer les réalisations d'une variable aléatoire, on peut utiliser la méthode d'inversion.
Cette méthode fonctionne si la fonction de répartition de la loi que nous cherchons à simuler est inversible. L'inverse de la fonction de répartition est appelée fonction quantile.

Définition : Soit F_X la fonction de répartition d'une variable aléatoire X. La fonction quantile de F_X est définie par :

$$F_X^{-1}(p) = \inf \{t \in \mathbb{R}, F_X(t) \geqslant p\} \tag{2.1}$$

Or si $U \sim \mathcal{U}(0,1)$ alors $F_X^{-1}(U)$ a pour fonction de répartition F_X
Ainsi, si la fonction quantile de la loi est connue, cette méthode permet de simuler des variables aléatoires qui ont la même loi que X.

En pratique, cette méthode est efficace pour générer des distributions discrètes.
Voici ci-dessous, l'algorithme de cette méthode qui est utilisé dans le cadre de ce mémoire.

2.2.2 Algorithme de la méthode d'inversion

Considérons une variable aléatoire X, de loi discète $(p_n)_{n\geq_0}$ que nous souhaitons simuler.
L'intervalle $[0;1]$ est découpé en sous intervalles dont les bornes sont les valeurs de la fonction de répartition de X c'est à dire les $\sum p_i$ croissantes.
Nous tirons un nombre U selon une loi uniforme $\mathcal{U}(0,1)$
L'indice de la borne supérieure de l'intervalle créé dans lequel se trouve U donne la valeur simulée de la distribution.
Si nous définissons la variable aléatoire N par :

$$\begin{cases} N = 0 \text{ si } U < p_0 \\ N = 1 \text{ si } p_0 \leq U < p_0 + p_1 \\ \dots \\ N = j \text{ si } \sum_{i=0}^{j-1} p_i \leq U < \sum_{i=0}^{j} p_i \\ \dots \end{cases} \tag{2.2}$$

Alors, d'après ce que nous avons dis précédemment, N a la même loi que X

2.2.3 Simulation des entrées en arrêt travail

Nous disposons d'une base d'actifs français par âge et sexe (voir figure 1.3). C'est à partir de cette base, que nous allons simuler les survenances d'arrêt de travail à l'aide de la table d'incidence Prim'act (voir figure 1.7) et de la méthode d'incidence.

La table d'incidence nous donne les taux q_x d'entrée en incapacité à partir de 3 mois d'ancienneté (90 jours de franchise) pour un individu d'âge x en fonction de son sexe.
Ainsi, à partir de la base, nous utilisons l'algorithme précédent :

Pour un âge x, un nombre U correspondant au nombre d'hommes actifs d'âge x (respectivement de femmes actives d'âge x) est tiré selon une loi uniforme $\mathcal{U}(0,1)$

Si $0 \leq U < q_x$ alors $N = 1$. Autrement dit, l'individu est entré en arrêt de travail à l'âge x
Si $q_x \leq U < 1$ alors $N = 0$. Autrement dit, l'individu est toujours actif à l'âge x

Nous répétons cet algorithme pour chaque âge (de 18 à 64 ans) en fonction du sexe de l'individu.
Nous obtenons ainsi le nombre total d'individus rentrés en incapacité pour chaque âge x.

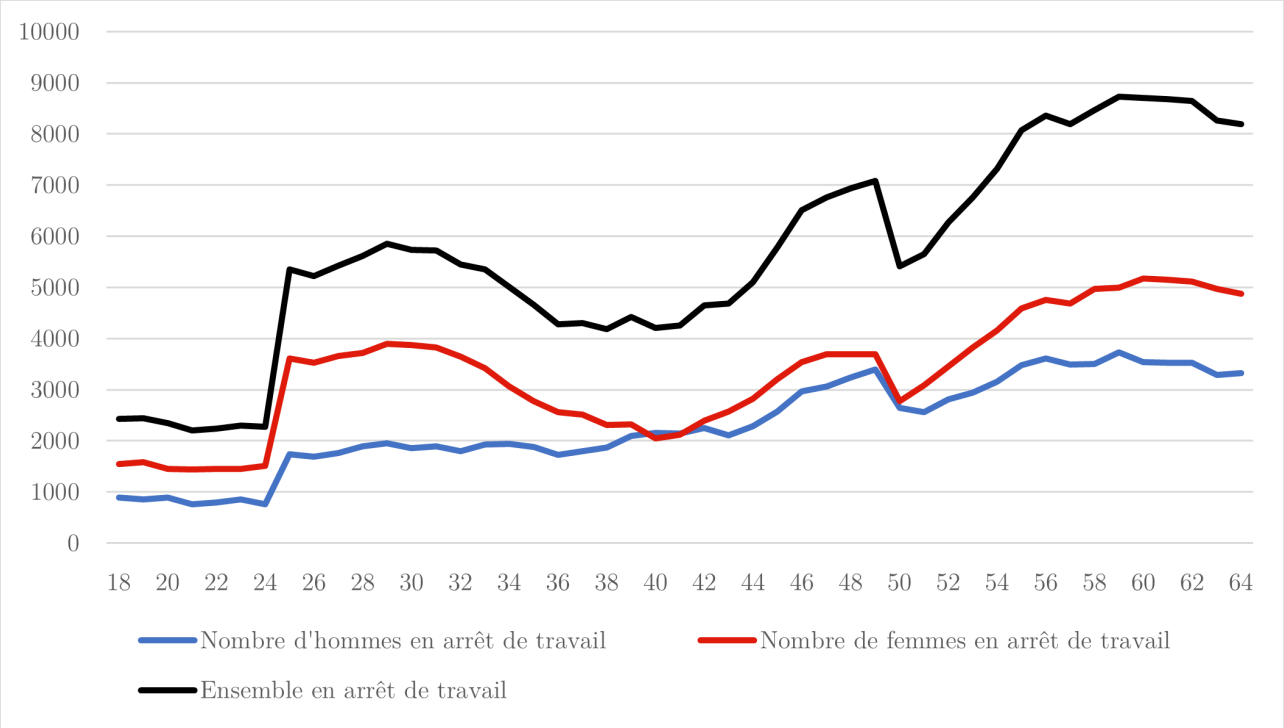


FIGURE 2.1 : Ensemble des individus (en milliers) entrés en arrêt de travail (avec 3 mois d'ancienneté)

Cette simulation nous permet donc d'obtenir le nombre de sinistres d'arrêt de travail.
Pour chaque sinistre, il est donc nécessaire maintenant de simuler les caractéristiques qui le constituent.

2.3 Simulation des caractéristiques du sinistre d'arrêt de travail

2.3.1 Description des composantes d'un sinistre

Un sinistre est caractérisé par différentes composantes que l'on retrouve dans le schéma suivant :

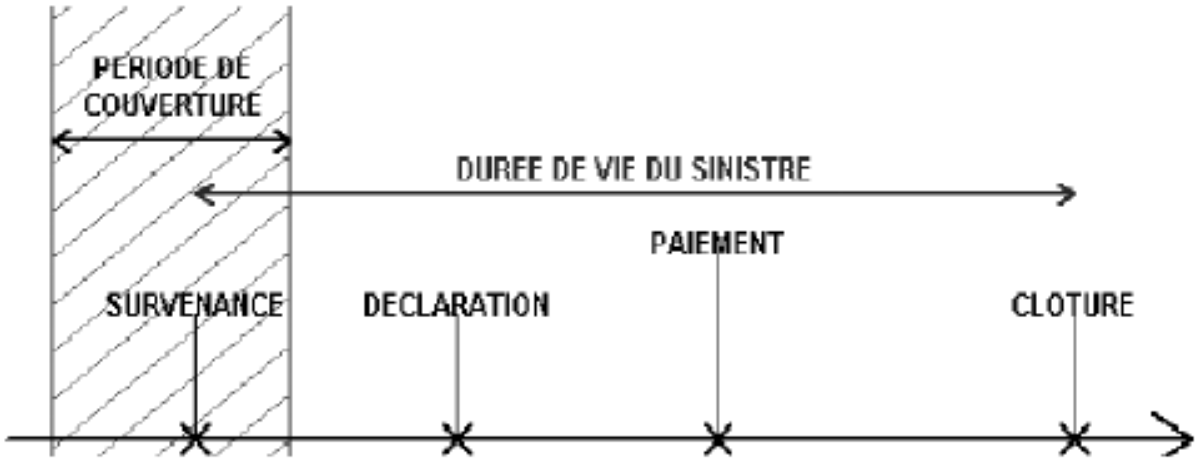


FIGURE 2.2 : Etapes de la vie d'un sinistre

- **La date de survenance du sinistre** : C'est la date qui permet à l'assureur de connaître le moment de la survenance d'un sinistre afin d'appliquer la garantie. Elle correspond généralement à la date de début d'indemnisation.

- **La date de déclaration du sinistre** : C'est la date à laquelle l'assureur prend conscience de la survenance du sinistre par l'assuré. Selon l'article L113-2 du code des assurances : "l'assuré est obligé de donner avis à l'assureur dès qu'il en a eu connaissance et au plus tard dans le délai fixé par le contrat, de tout sinistre de nature à entraîner la garantie de l'assureur. Le délai ne peut être inférieur à cinq jours ouvrés".

- **Les paiements** : Il s'agit de l'indemnisation contractuelle que l'assureur verse auprès de son assuré. Les paiements peuvent prendre plusieurs formes selon les contrats (rentes, capital, etc ...). Dans le cadre de ce mémoire sur l'arrêt de travail, nous supposons que la prestation est une rente de 1 euro versé par jour restée en incapacité (après un délai de 90 jours de franchise). Dans la suite du mémoire, les montants s'exprimeront en nombre de mois indemnisés.

- **La date de clôture** : C'est la date à laquelle l'assuré a été entièrement indemnisé dans les termes du contrat. Dans le cadre de ce mémoire, la date de clôture correspond donc à la date à laquelle les jours passés en état d'incapacité ont été entièrement indemnisés.

2.3.2 Simulation des caractéristiques du sinistre d'arrêt de travail

Grâce à la simulation précédente, nous avons obtenu un nombre d'individus par âge rentrés en incapacité (avec une ancienneté de 3 mois). Pour chaque sinistre correspondant, nous allons donc simuler 3 caractéristiques :

- La date de survenance ;
- La date de déclaration ;
- La durée de maintien en incapacité

Date de survenance

Nous simulons la date de survenance de chaque sinistre uniformément dans l'année N.
Ainsi, pour chaque entrée en incapacité, nous tirons un nombre de jours U selon une loi uniforme $\mathcal{U}(1, 365)$.
Ce nombre est ensuite ajouté à la date "31/12/N-1" pour obtenir la date de survenance du sinistre de l'année N.

Les sinistres de la base de simulation de l'année N sont bien uniformément répartis.

Date de déclaration

Nous simulons le délai de déclaration (en jours) selon une loi log-normale. Le choix de cette loi est justifié par le caractère positif de son support et le fait qu'elle soit simple à mettre en pratique.

Les paramètres de la loi log-normale sont choisis de sorte à obtenir une moyenne d'environ 2 mois (soit environ 60 jours) et un quantile à 90% égal environ à 3 mois (soit 90 jours).

Minimum	**1er Quartile**	**Mediane**	**Moyenne**	**3ème Quartile**	**Quantile à** 90%	**Maximum**
13	48	60	64	76	93	267

FIGURE 2.3 : Statistiques descriptives du délai de déclaration

Durée de maintien en incapacité

Nous simulons les durées de maintien en incapacité de chaque individu en arrêt de travail (avec 3 mois d'ancienneté) en utilisant la méthode d'inversion avec la loi donnée par la table du BCAC.

Les 3 premières colonnes de cette table sont retirées car les individus ont déjà 3 mois d'ancienneté. Le but est donc de simuler les durées restantes en terme d'indemnisation (donc de 1 à 33 mois).

A partir de cette table, nous reconstruisons les $q_{x,t}$ (probabilité de sortie de l'état d'incapacité d'un individu d'âge x entre le mois t et $t+1$) avec la formule (1.11).

La fonction de survie de la loi de maintien pour un individu d'âge x est $S_x(t) = \prod_{y|y<t} (1 - q_{x,y})$

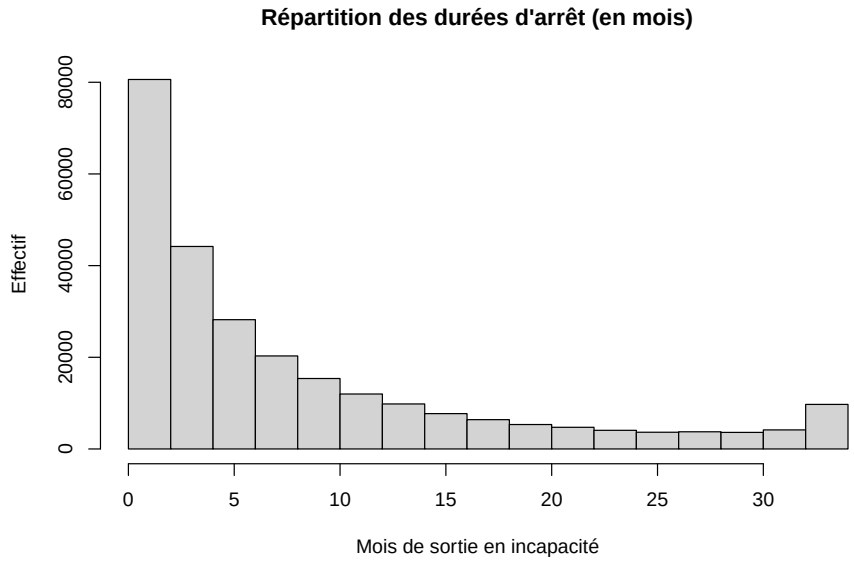
Nous appliquons ensuite l'algorithme de la méthode d'inversion :

L'intervalle $[0, 1]$ est subdivisé en sous intervalles dont les bornes sont les termes de la fonction de survie : $[0,\ \mathrm{S}(32)[; [\mathrm{S}(32), \mathrm{S}(31)[; \ldots ; [\mathrm{S}(2), \mathrm{S}(1)[; [\mathrm{S}(1), 0]$.
Pour un âge x, nous tirons U (nombre d'individus entrés en incapacité à l'âge x) selon une loi uniforme $\mathcal{U}(0, 1)$.

Si $0 \leq \mathrm{U} < \mathrm{S}(1)$ alors l'individu d'âge x sort de l'incapacité le 1er mois d'indemnisation
Si $\mathrm{S}(1) \leq \mathrm{U} < \mathrm{S}(2)$ alors l'individu d'âge x sort de l'incapacité le 2ème mois d'indemnisation
$\vdots$
Si $\mathrm{S}(31) \leq \mathrm{U} < \mathrm{S}(32)$ alors l'individu d'âge x sort de l'incapacité le 32ème mois d'indemnisation
Si $\mathrm{S}(32) \leq \mathrm{U} \leq 1$ alors l'individu d'âge x sort de l'incapacité le 33ème mois d'indemnisation

Afin d'obtenir une durée précise en jours passés en incapacité jusqu'au mois i, on tire une loi uniforme $\mathcal{U}(30 \times (i-1), 30 \times i)$. Ce qui permet , en ajoutant le nombre de jours passé en incapacité avec la date de survenance, d'en déduire la date de sortie.

Ainsi, pour chaque individu rentré en incapacité à l'âge x après 3 mois d'ancienneté, nous obtenons leur durée de maintien dans cet état.



Nous observons que le nombre de mois de sortie en incapacité décroit quand l'ancienneté augmente. Il y a un pic au niveau du 33ème mois car, l'incapacité durant 3 ans, les individus sortent de cet état. Les individus qui ont une ancienneté importante ont également une probabilité plus forte à demeurer en incapacité jusqu'à la fin ce qui explique également ce pic.

Conclusion

Nous avons donc simuler à partir d'une structure par âge des actifs français, une base contenant tous les individus en arrêt de travail avec la date de survenance du sinistre, la date de déclaration et la durée de maintien de l'individu en incapacité.

A partir de cette base, nous pouvons donc tester les différentes méthodes de provisionnement utilisées par les compagnies d'assurance afin de comparer les provisions estimées aux durées réelles simulées. C'est ce que nous allons voir dans le chapitre suivant.

Chapitre 3

Méthodes de provisionnement

3.1 Les provisions en assurance non vie

3.1.1 L'importance du provisionnement

L'assurance est une activité économique singulière dont la spécificité est l'inversion du cycle de production. En effet, l'organisme assureur reçoit une prime de la part de l'assuré et s'engage, en contrepartie, à lui verser une prestation en cas de réalisation d'un risque aléatoire prévu dans le contrat.

Ainsi, l'enjeu pour une compagnie d'assurance est d'estimer le montant des réserves à constituer afin d'être capable de faire face à toutes les indemnisations qu'elle aura à verser pour les sinistres qui seront survenus pendant la période couverte par la police d'assurance.

Il est donc primordial pour l'assureur de garder une certaine somme d'argent pour honorer ses engagements envers l'assuré en cas de réalisation du sinistre. C'est ce que l'on appelle les provisions.

Elles représentent une proportion importante du passif du bilan d'une compagnie d'assurance (voir figure ci-après). L'enjeu du provisionnement est donc d'estimer les prestations futures en fonction du caractère aléatoire des sinistres.

Les provisions se définissent différemment selon la norme considérée : norme comptable française ou norme solvabilité II.

Norme comptable

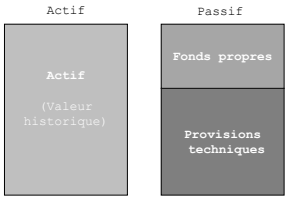


FIGURE 3.1 : Bilan comptable simplifié d'une compagnie d'assurance

En norme comptable, l'objectif est d'honorer les engagements pris à l'égard des assurés. Ainsi les actifs sont comptabilisés en valeur historique. Les provisions techniques sont évaluées à l'aide d'hypothèses prudentes.

Norme Solvabilité II

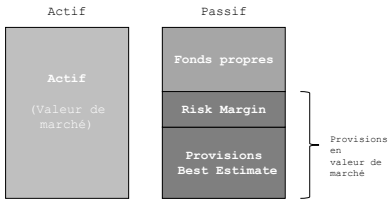


FIGURE 3.2 : Bilan solvabilité II simplifié d'une compagnie d'assurance

En norme Solvabilité II, la vision du bilan devient économique. En effet, les actifs sont comptabilisés à la juste valeur (ou valeur de marché) et non plus en valeur comptable.

La juste valeur est « le prix qui serait reçu pour la vente d'un actif ou payé pour le transfert d'un passif lors d'une transaction normale entre des intervenants du marché à la date d'évaluation » (définition IFRS 13).

Les provisions techniques du passif sont également comptabilisées à la juste valeur. Elles sont composées de la Risk Margin et du Best Estimate.

La **Risk Margin** « est calculée de manière à garantir que la valeur des provisions techniques est équivalente au montant que les entreprises d'assurance et de réassurance demanderaient pour reprendre et honorer les engagements d'assurance et de réassurance ».

Le **Best Estimate** est « la moyenne pondérée en fonction de leur probabilité des futurs flux de trésorerie compte tenu de la valeur temporelle de l'argent, laquelle est estimée sur la base de la courbe des taux sans risque pertinente. » (Directive Solvabilité 2 art 77).

3.1.2 La PSAP

La PSAP désigne les provisions pour sinistres à payer. Selon l'article R331-17 du Code des Assurances "la PSAP est la valeur estimative des dépenses en principal et en frais, internes nécessaires au règlement de tous les sinistres survenus et non payés".

Les provisions pour sinistres à payer sont souvent les provisions les plus importantes en montant. C'est pourquoi elles nécessitent une évaluation très précise. L'incertitude peut donc porter sur la date de déclaration ainsi que sur le coût final du sinistre.

Afin d'estimer les provisions pour régler les prestations futures, l'assureur se place à une date de calcul au 31/12/N
D'après la description de la vie d'un sinistre que nous avons vue dans la partie précédente, on peut distinguer plusieurs types de sinistres :

Les sinistres survenus et connus : il s'agit des sinistres qui sont survenus avant la date de calcul et dont la date de déclaration est antérieure au 31/12/N. Parmi ces sinistres, nous pouvons distinguer les sinistres clôturés et les sinistres en cours.
Les sinistres clôturés sont les sinistres dont la date de sortie est antérieure à la date de calcul. Pour ces sinistres, nous connaissons donc leur prestation.
Les sinistres en cours sont les sinistres dont on connait une partie des prestations à régler mais une autre partie de ces prestations reste indeterminée car la date de sortie au 31/12/N est inconnue. Il est donc nécessaire d'estimer une provision pour ces sinistres.

Les sinistres survenus et non connus : il s'agit des sinistres survenus avant la date de calcul mais dont la date de déclaration est postérieure au 31/12/N, c'est à dire que l'assureur n'a pas encore pris connaissance du sinistre. C'est ce que l'on appelle les sinistres tardifs ou IBNR (*Incurred But Not Reported*) qui nécessiteront une estimation à part.
Le but de ce mémoire est d'utiliser et de comparer les méthodes de provisionnement usuelles sur un grand nombre de simulations. Nous allons utiliser 3 méthodes de provisionnements :

- La méthode de provisionnement ligne à ligne pour estimer le montant des sinistres en cours ;
- La méthode de provisionnement agrégée : la méthode de Chain Ladder sur un triangle en montant et en nombre de sinistres afin d'estimer la PSAP et déduire les sinistres IBNR ;
- Une méthode de provisionnement à l'aide de la loi d'incidence afin d'estimer un montant d'IBNR.

3.2 Méthode de provisionnement ligne à ligne

3.2.1 Elaboration de la base du point de vue de l'assureur au 31/12/N

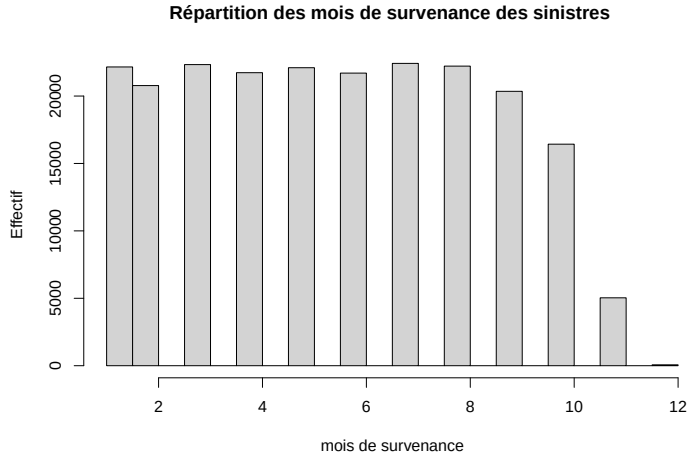
A partir de la simulation réalisée au chapitre 2, nous élaborons une base dans laquelle se trouve les sinistres survenus et connus de l'assureur. Nous nous plaçoons à la date de calcul du 31/12/N.

Autrement dit, cette base contient les informations du point de vue de l'assureur à cette date. Elle comporte donc les sinistres survenus durant l'année N et dont la date de déclaration est antérieure au 31/12/N, c'est à dire les sinistres survenus et connus au 31/12/N.

De plus l'assureur, à cette date, ne connait pas la durée totale d'arrêt des sinistres en cours. Ainsi, pour les individus qui ne sont pas encore sortis au 31/12/N, nous avons censuré les durées d'arrêt à cette date.

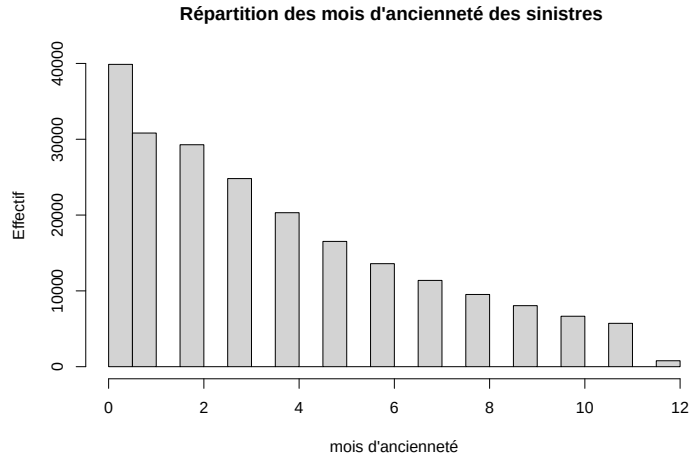
Cette base contient les variables suivantes : "âge à la survenance", "Date de survenance", "Date de déclaration", "Clos ou en cours", "Durée de maintien", "Ancienneté".

- La variable "clos ou en cours" est une indicatrice égale à 0 si le sinistre est clos et 1 si le sinistres est en cours.
- La variable "durée de maintien" est la durée de maintien (en jours) de l'individu en incapacité à la date de calcul.
- La variable "Ancienneté" désigne l'ancienneté (en mois) de l'individu en incapacité à la date de calcul. Par exemple, si la durée de maintien est comprise entre 0 et 30 jours l'ancienneté est 0.



Nous observons que les mois de survenance de cette base sont uniformément répartis jusqu'au 9ème mois. A partir du 10ème mois, les survenances diminuent. Il s'agit de l'absence des sinistres survenus qui n'ont pas été déclarés à l'assureur.

En effet, puisque la durée de déclaration était en moyenne de 2 mois : au bout du 10ème mois de l'année N, beaucoup de sinistres ne sont pas déclarés et sont donc inconnus aux yeux de l'assureur, ce qui explique cette chute.



Nous rappelons que l'ancienneté correspond à la durée d'indemnisation de l'incapacité d'un individu depuis la survenance (il s'agit de la durée d'indemnisation et non de la durée totale de maintien en incapacité , c'est à dire qu'une ancienneté de 0 depuis la survenance correspond à une durée de maintien en incapacité de 3 mois en comptant le délai de franchise).
Ce graphique représente la répartition de l'ancienneté des sinistres de la base.
En résumé, nous nous retrouvons avec une base sur une année N, de sinistres survenus et connus au 31/12/N. Parmi ces sinistres, certains sont clôturés, c'est à dire que leur date de sortie est antérieure au 31/12/N ; d'autres sont en cours et ne sont pas terminés.

Nous allons donc dans un premier temps, utiliser une méthode de provisionnement ligne à ligne afin d'estimer les provisions pour les sinistres en cours. Nous estimerons dans un second temps les sinistres tardifs (IBNR) qui ne sont pas connus au 31/12/N à l'aide des méthodes agrégées.

3.2.2 Méthode de provisionnement ligne à ligne

Afin d'estimer la provision des sinistres en cours, nous nous basons sur une approche ligne à ligne. C'est à dire que nous estimons pour chaque sinistre en cours, la durée restante en incapacité en fonction de l'âge à la survenance et de l'ancienneté.

Soit un individu d'âge x à la survenance et d'ancienneté y. Le coefficient de provisionnement est donné par la formule suivante :

$$\mathrm{PM}_{\mathrm{y}}^{\mathrm{x}} = \frac{1}{\mathrm{L}_{\mathrm{x,y}}} \sum_{\mathrm{k=y+1}}^{35} \mathrm{L}_{\mathrm{x,k}} + \frac{1}{2} \tag{3.1}$$

Nous obtenons ainsi pour chaque sinistre en cours, une estimation de la durée restante (en mois) dans l'état d'incapacité. Ce sont les provisions pour les sinistres en cours.

Lorsque nous sommons ces provisions avec les durées d'ancienneté des sinistres clos et en cours, nous obtenons une estimation de la charge des sinistres survenus et connus de l'année N.

Nous comparons l'estimation de la charge ultime des sinistres survenus et connus à la charge ultime simulée (en retirant les sinistres non connus) :

Charge Ultime estimé (hors IBNR)	**Charge Ultime simulée (hors IBNR)**	**Ecart en pourcentage**
1 655 864	1 651 673	0,25%

FIGURE 3.3 : Résultat de l'estimation de la charge ultime des sinistres de l'année N

Nous observons qu'en ce qui concerne cette simulation, cette méthode sur provisionne légèrement les prestations mais reste très précise.

Afin d'estimer le montant d'IBNR, nous allons utiliser les méthodes de provisionnement agrégée (à base de triangle) qui donneront une estimation de la charge ultime en prenant en compte les IBNR. C'est ce qui concerne la prochaine partie.

3.3 Méthode de provisionnement agrégée : la méthode de Chain Ladder

La méthode Chain Ladder est une méthode de provisionnement agrégée à la différence de la méthode ligne à ligne utilisée précédemment. C'est une méthode déterministe qui est basée sur la construction de triangles de paiements afin de modéliser les coûts des sinistres.

3.3.1 Triangulation des coûts des sinistres

Les coûts de sinistres sont modélisés à l'aide de triangle de développement, ils sont agrégés par date de survenance.
On utilisera les notations suivantes :

- i le mois de survenance du sinistre, $i \in [0, ..., n]$;
- j le mois de développement $j \in [0, ..., n]$;
- $X_{i,j}$ le montant des sinistres survenus le mois i et payé après le délai de j mois de developpement
- $C_{i,j}$ le montant cumulé des sinistres survenus le mois i et payé après le délai de j mois de developpement : $C_{i,j} = \sum_{k=0}^{j} X_{i,k}$
- $S_i = C_{i,n}$ les charges ultimes correspondant aux sinistres survenus le mois i $= 0, \ldots, n$
- R_i les provisions par mois de survenance avec $R_0 = 0$ et $R_i = \sum_{k=n-i+1}^{n} X_{i,k}$ pour $i = 1, \ldots, n$
- R le montant total des provisions : $R = \sum_{k=0}^{n} R_i$

	0	1	...	j	...	$n-i$	...	$n-1$	n
0	$C_{0,1}$	$C_{0,1}$	...	$C_{0,j}$	...	...	...	$C_{0,n-1}$	$C_{0,n}$
1	$C_{1,0}$	$C_{1,1}$	...	$C_{1,j}$	...	...	...	$C_{1,n-1}$	
...	...	...	...	...	...	...	...		
i	...	...	...	$C_{i,j}$	...	$C_{i,n-i}$			
...	...	...	...	...	...				
$n-j$	...	...	...	$C_{n-j,j}$					
...	...	...	...						
$n-1$	$C_{n-1,0}$	$C_{n-1,1}$	...						
n	$C_{n,0}$								

FIGURE 3.4 : Triangle de paiements cumulés

Ce triangle contient les paiements connus de l'assureur, il s'agit du triangle supérieur noté : $T_S = \{(C_{i,j}) \mid 0 \leq i + j \leq n\}$.
Le but des méthodes de provisionnement agrégées est d'estimer la partie inférieure du triangle $T_I = \left\{\left(\hat{C}_{i,j}\right) \mid i + j > n; i \leq n; j \leq n\right\}$.
Nous allons utiliser la méthode de Chain Ladder qui est la plus utilisée en pratique afin d'estimer ces charges futures.

3.3.2 La méthode de Chain Ladder

Si on considère le mois de survenance i dont on décrit ci-dessous le développement complet en paiements cumulés au 31/12/N :

	0	...	n − i	...	j	j + 1	...	n
i	$C_{i,0}$	...	$C_{i,n-i}$	...	$C_{i,j}$	$C_{i,j+1}$	...	$C_{i,n}$

Le j-ème facteur de développement est $f_j = \frac{C_{i,j+1}}{C_{i,j}}$

La méthode de Chain Ladder est une méthode déterministe fréquemment utilisée car facile à mettre en oeuvre. Elle s'applique à des triangles de paiements cumulés ou des triangles de charges.
Elle est basée sur l'utilisation des facteurs de développement f_j sous 2 hypothèses :

- (H1) Les paiements cumulés $C_{i,j}$ sont indépendants des années de survenance i.
- (H2) II existe des facteurs de développements $f_0, \ldots, f_{j-1}$ tel que $\forall 0 \leq j \leq n-1$,

$$\forall 0 \leq i \leq n \text{ on a } C_{i,j+1} = f_j C_{i,j}$$

Les facteurs de développement de Chain Ladder f_j sont estimés de la manière suivante :

$$\widehat{f}_j = \frac{\sum_{i=0}^{n-j-1} c_{i,j+1}}{\sum_{k=0}^{n-j-1} c_{i,j}} \tag{3.2}$$

Cette estimation est un rapport de la moyenne des paiements cumulés du mois de développement $j+1$ sur la moyenne des paiements cumulés du mois de développement j

A l'aide de ces facteurs, nous pouvons donc estimer :

- les charges ultimes par exercice de survenance $\widehat{S_i} = C_{i,n-i} \prod_{k=n-i}^{n-1} \widehat{f_k}$
- les provisions par exercice de survenance $\widehat{R_{\text{i}}} = \widehat{C_{\text{i},n}} - C_{i,n-i}$
- la provision totale $\hat{R} = \sum_{i=0}^{n} \widehat{R}_i$

3.3.3 Mise en application : triangle en nombre de sinistres

Nous rappelons que nous disposons d'une base contenant tous les sinistres d'arrêt de travail survenus et connus l'année N. L'objectif est d'estimer la charge ultime que représentent tous ces sinistres ainsi que les IBNR à l'aide de la méthode de Chain Ladder.

Afin de construire le triangle de paiements, nous simulons un historique de l'année $N-2$ et $N-1$ exactement de la même façon que l'année N. Les sinistres des années $N-2$ et $N-1$ sont tous connus au 31/12/N. Les durées d'arrêts sont censurées au 31/12/N.

Ainsi, en concaténant les bases des années $N-2$ et $N-1$ avec la base de l'assureur de l'année N, nous disposons des arrêts dont les dates de survenance sont comprises entre le 01/01/N-2 et le 31/12/N.

Cet historique, nous permet de réaliser le triangle en nombre de sinistres de la manière suivante :

- En ligne se trouve le nombre de sinistres par mois de survenance de janvier de l'année N-2 à décembre de l'année N ;
- En colonne se trouve le nombre de sinistres par mois de déclaration

Le but de la méthode de Chain Ladder est d'estimer un nombre de sinistres de l'année N à partir de l'historique, à partir duquel nous déduisons le nombre d'IBNR. Le montant de la charge ultime (en nombre de mois d'indemnisation) est calculé en multipliant le nombre de sinistres par un coût moyen.

Le coût moyen que nous considérons est la moyenne des durées (en nombre de mois) d'indemnisation des sinistres clos de tout l'historique (du 01/01/N-2 au 31/12/N) qui est de $7,62$ mois.
Nous obtenons le triangle cumulé suivant :

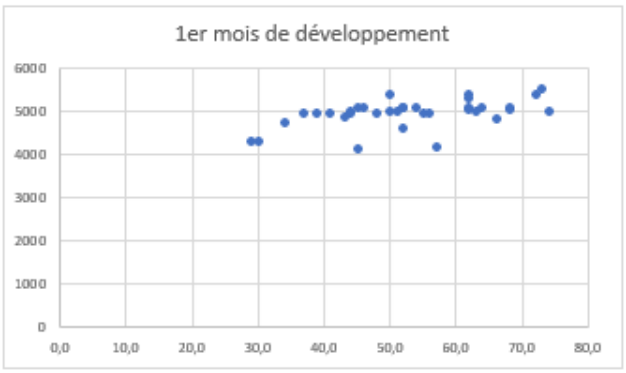
	0	1	2	3	4	5	6	7	8	9
N-2-01	45	4144	15756	20802	21938	22125	22160	22169	22169	22169
N-2-02	30	4333	14834	19382	20266	20436	20464	20472	20473	20474
N-2-03	52	5085	16712	21394	22399	22588	22618	22621	22621	22621
N-2-04	51	5012	15960	20591	21512	21666	21696	21703	21703	21703
N-2-05	62	5044	16460	21309	22290	22466	22491	22496	22496	22496
N-2-06	48	4963	16309	20671	21572	21733	21761	21763	21763	21763
N-2-07	72	5406	16607	21319	22185	22390	22409	22414	22415	22415
N-2-08	64	5101	16588	21283	22279	22432	22454	22459	22461	22461
N-2-09	37	4965	15947	20665	21621	21761	21792	21792	21792	21793
N-2-10	74	5000	16565	21256	22188	22369	22396	22402	22406	22406
N-2-11	54	5091	16333	20639	21678	21862	21902	21910	21911	21911
N-2-12	50	5393	15739	20791	21795	21953	21980	21994	21996	21996
N-1-01	57	4180	15659	20691	21855	22045	22085	22094	22095	22095
N-1-02	29	4319	14724	19239	20126	20283	20300	20305	20305	20305
N-1-03	43	4902	16545	21299	22276	22462	22490	22493	22494	22495
N-1-04	45	5097	16032	20743	21691	21819	21846	21853	21855	21856
N-1-05	63	5024	16844	21595	22550	22725	22751	22753	22754	22754
N-1-06	44	5015	16289	20696	21629	21778	21808	21814	21814	21814
N-1-07	62	5391	16602	21299	22237	22402	22425	22426	22429	22429
N-1-08	52	5107	16691	21242	22259	22411	22439	22444	22444	22444
N-1-09	46	5097	15796	20389	21308	21441	21477	21481	21481	21481
N-1-10	56	4952	16476	21287	22172	22365	22390	22393	22393	22393
N-1-11	41	4964	16079	20328	21302	21462	21499	21501	21501	21501
N-1-12	62	5311	15991	20799	21803	21980	22008	22012	22014	22015
N-01	52	4618	16049	20868	21927	22125	22147	22156	22157	22157
N-02	34	4731	15149	19696	20560	20736	20762	20766	20766	20766
N-03	68	5068	16538	21131	22182	22339	22359	22365	22365	22365
N-04	44	4992	15968	20676	21581	21708	21733	21733	21734	0
N-05	66	4851	16197	20972	21931	22104	22124	22128	0	0
N-06	62	5085	16223	20589	21520	21675	21698	0	0	0
N-07	73	5545	16596	21338	22296	22442	0	0	0	0
N-08	68	5087	16553	21200	22195	0	0	0	0	0
N-09	39	4951	15690	20370	0	0	0	0	0	0
N-10	55	4989	16395	0	0	0	0	0	0	0
N-11	50	5015	0	0	0	0	0	0	0	0
N-12	68	0	0	0	0	0	0	0	0	0

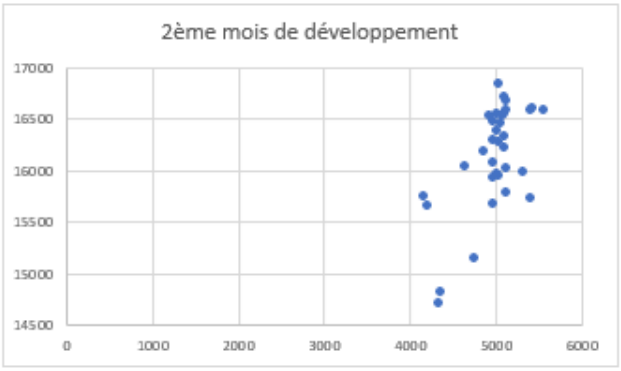
FIGURE 3.5 : Triangle cumulé en nombre de sinistres

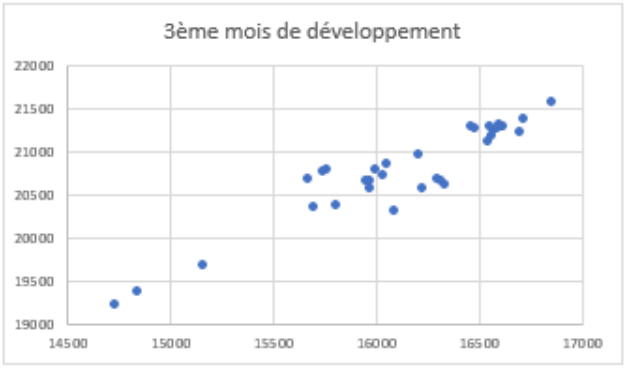
Nous constatons que très peu de sinistres sont déclarés à partir du mois de la survenance et la majorité sont déclarés à partir du 2ème mois. Cet écart est justifié par le choix de la simulation avec une loi log-normale de moyenne 60 jours.

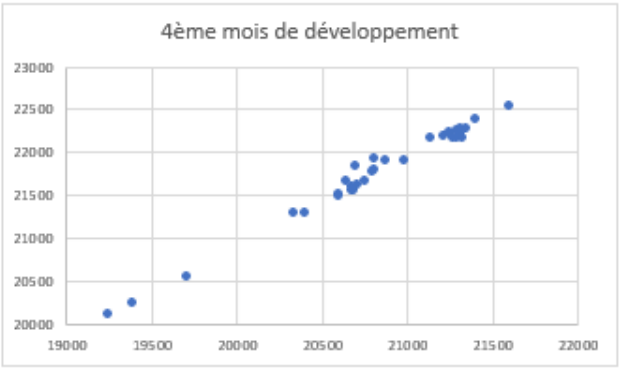
Validation des hypothèses de Chain Ladder :

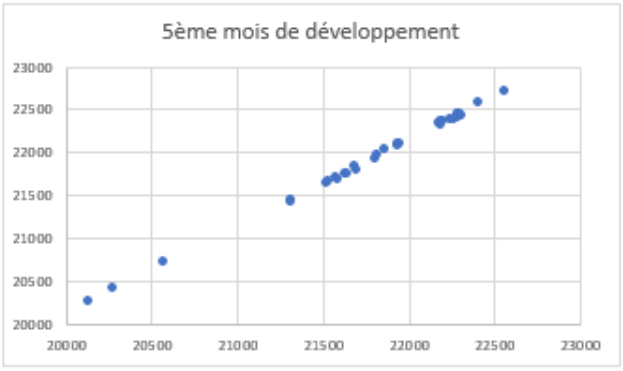
Afin de valider l'hypothèse de Chain Ladder, nous traçons les nuages de points $\forall i, (C_{i,j}, C_{i,j+1})$ avec j allant jusqu'à 6. Les mois suivants de développement présentent le même alignement que le graphique du mois 6 et ne sont donc pas représentés.

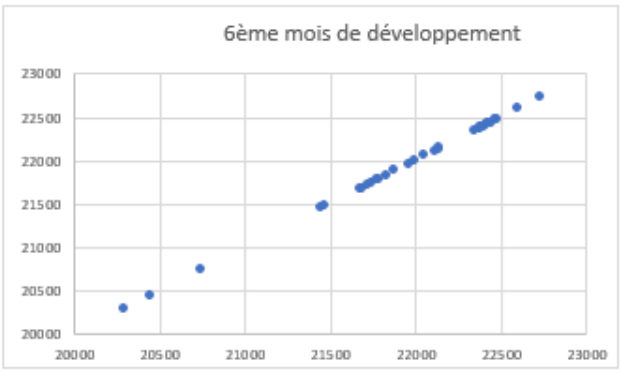












Les deux premiers mois de développement présentent naturellement une structure assez irrégulière. En effet, la plupart des sinistres sont déclarés au bout du 2 ème mois.
A partir du 3ème mois, les graphiques présentent des points alignés sur une droite passant par l'origine ce qui permet de valider l'utilisation de la méthode de Chain Ladder.

En utilisant la méthode de Chain Ladder sur ce triangle cumulé, nous obtenons en dernière colonne le nombre ultime de sinistres par mois de survenance.

	0	1	2	3	4	5	6	7	8	9
2018-01	45,0	4144	15756	20802	21938	22125	22160	22169	22169	22169
2018-02	30,0	4333	14834	19382	20266	20436	20464	20472	20473	20474
2018-03	52,0	5085	16712	21394	22399	22588	22618	22621	22621	22621
2018-04	51,0	5012	15960	20591	21512	21666	21696	21703	21703	21703
2018-05	62,0	5044	16460	21309	22290	22466	22491	22496	22496	22496
2018-06	48,0	4963	16309	20671	21572	21733	21761	21763	21763	21763
2018-07	72,0	5406	16607	21319	22185	22390	22409	22414	22415	22415
2018-08	64,0	5101	16588	21283	22279	22432	22454	22459	22461	22461
2018-09	37,0	4965	15947	20665	21621	21761	21792	21792	21792	21793
2018-10	74,0	5000	16565	21256	22188	22369	22396	22402	22406	22406
2018-11	54,0	5091	16333	20639	21678	21862	21902	21910	21911	21911
2018-12	50,0	5393	15739	20791	21795	21953	21980	21994	21996	21996
2019-01	57,0	4180	15659	20691	21855	22045	22085	22094	22095	22095
2019-02	29,0	4319	14724	19239	20126	20283	20300	20305	20305	20305
2019-03	43,0	4902	16545	21299	22276	22462	22490	22493	22494	22495
2019-04	45,0	5097	16032	20743	21691	21819	21846	21853	21855	21856
2019-05	63,0	5024	16844	21595	22550	22725	22751	22753	22754	22754
2019-06	44,0	5015	16289	20696	21629	21778	21808	21814	21814	21814
2019-07	62,0	5391	16602	21299	22237	22402	22425	22426	22429	22429
2019-08	52,0	5107	16691	21242	22259	22411	22439	22444	22444	22444
2019-09	46,0	5097	15796	20389	21308	21441	21477	21481	21481	21481
2019-10	56,0	4952	16476	21287	22172	22365	22390	22393	22393	22393
2019-11	41,0	4964	16079	20328	21302	21462	21499	21501	21501	21501
2019-12	62,0	5311	15991	20799	21803	21980	22008	22012	22014	22015
2020-01	52,0	4618	16049	20868	21927	22125	22147	22156	22157	22157
2020-02	34,0	4731	15149	19696	20560	20736	20762	20766	20766	20766
2020-03	68,0	5068	16538	21131	22182	22339	22359	22365	22365	22365
2020-04	44,0	4992	15968	20676	21581	21708	21733	21733	21734	21734.2
2020-05	66,0	4851	16197	20972	21931	22104	22124	22128	22128.8	22129.0
2020-06	62,0	5085	16223	20589	21520	21675	21698	21702.9	21703.8	21703.9
2020-07	73,0	5545	16596	21338	22296	22442	22470.1	22475.2	22476.1	22476.3
2020-08	68,0	5087	16553	21200	22195	22364.5	22392.5	22397.6	22398.5	22398.7
2020-09	39,0	4951	15690	20370	21316.2	21478.9	21505.9	21510.8	21511.6	21511.8
2020-10	55,0	4989	16395	21137.9	22119.8	22288.7	22316.6	22321.7	22322.6	22322.8
2020-11	50,0	5015	16306.3	21023.5	22000.1	22168.1	22195.9	22200.9	22201.8	22201.9
2020-12	68,0	6389.4	20775.0	26785.0	28029.2	28243.3	28278.6	28285.1	28286.2	28286.4

FIGURE 3.6 : Triangle cumulé en nombre de sinistres complet

Nous observons que pour le dernier mois de l'année N, la méthode a tendance à surestimer le nombre de sinistres. Ceci peut s'expliquer à cause de la cumulation de l'incertitude des coefficients de passage.

En sommant les charges ultimes par mois de survenance de l'année N, nous obtenons une estimation du nombre total de sinistres pour cette année. Les IBNR sont déduits en soustrayant cette charge ultime avec la diagonale principale.

Nous obtenons les résultats suivants :

Nombre de sinistres ultime estimé	**Nombre de sinistres simulé**	**Ecart en pourcentage**
270 051	262 758	2,78%

FIGURE 3.7 : Résultat de l'estimation du nombre ultime des sinistres de l'année N

En multipliant ce nombre de sinistres avec le coût moyen, nous obtenons une estimation de la charge ultime pour l'année N :

Charge ultime de sinistres estimé	**Charge totale de sinistres simulée**	**Ecart en pourcentage**
2 060 153	2 002 932	2,85%

FIGURE 3.8 : Résultat de l'estimation de la charge totale des sinistres de l'année N

La méthode de Chain Ladder donne une estimation plutôt correcte de la charge ultime à indemniser pour l'année N.

On en déduit une estimation des IBNR (à partir du nombre ultime de sinistres) et de leur charge que l'on compare avec la simulation :

Nombre d'IBNR estimé	**Nombre d'IBNR simulé**	**Ecart en pourcentage**
52 718	45 425	16,05%

FIGURE 3.9 : Résultat de l'estimation du nombre d'IBNR de l'année N

Charge d'IBNR estimé	**Charge d'IBNR simulé**	**Ecart en pourcentage**
402 174	346 215	16,16%

FIGURE 3.10 : Résultat de l'estimation de la charge des IBNR de l'année N

Nous constatons que la méthode de Chain Ladder donne une estimation plutôt correcte de la charge globale de l'année N, mais surprovisionne les durées d'indemnisation des IBNR.

Cet écart peut s'expliquer à cause de l'accumulation de l'incertitude des coefficients de passage sur le dernier mois de survenance, ainsi que l'incertitude en elle même des coefficients de passage sur les premiers mois de développement.

Nous allons voir dans la prochaine section une autre méthode pour estimer ces tardifs basée sur la loi d'incidence.

3.4 Méthode de provisionnement des IBNR à l'aide de la loi d'incidence

Cette méthode consiste à estimer un nombre total de sinistres en arrêt de travail avec 3 mois d'ancienneté à l'aide de la loi d'incidence Prim'Act.

Pour cela, nous repartons de la structure par âge des actifs français (figure 1.3). Pour chaque âge, nous multiplions le nombre d'actifs (hommes et femmes) par le taux d'incidence donné par la table d'incidence Prim'Act.

Nous obtenons ainsi, pour chaque âge, une estimation du nombre d'individus entrés en arrêt de travail avec 3 mois d'ancienneté :

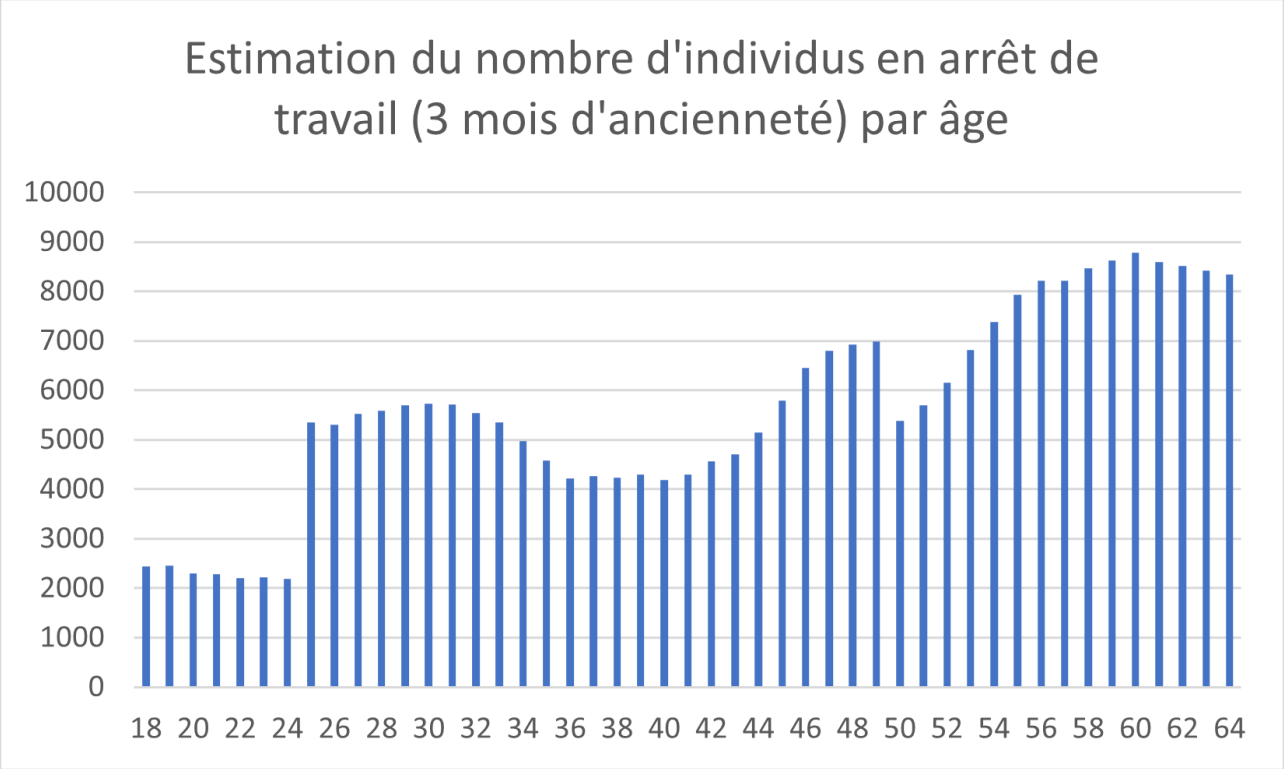


FIGURE 3.11 : Estimation du nombre de sinistres en arrêt de travail

Le nombre d'IBNR est ainsi déduit en soustrayant le nombre total estimé d'individus en arrêt de travail au nombre de sinistres de la base de l'assureur de l'année N. Nous obtenons les résultats suivants :

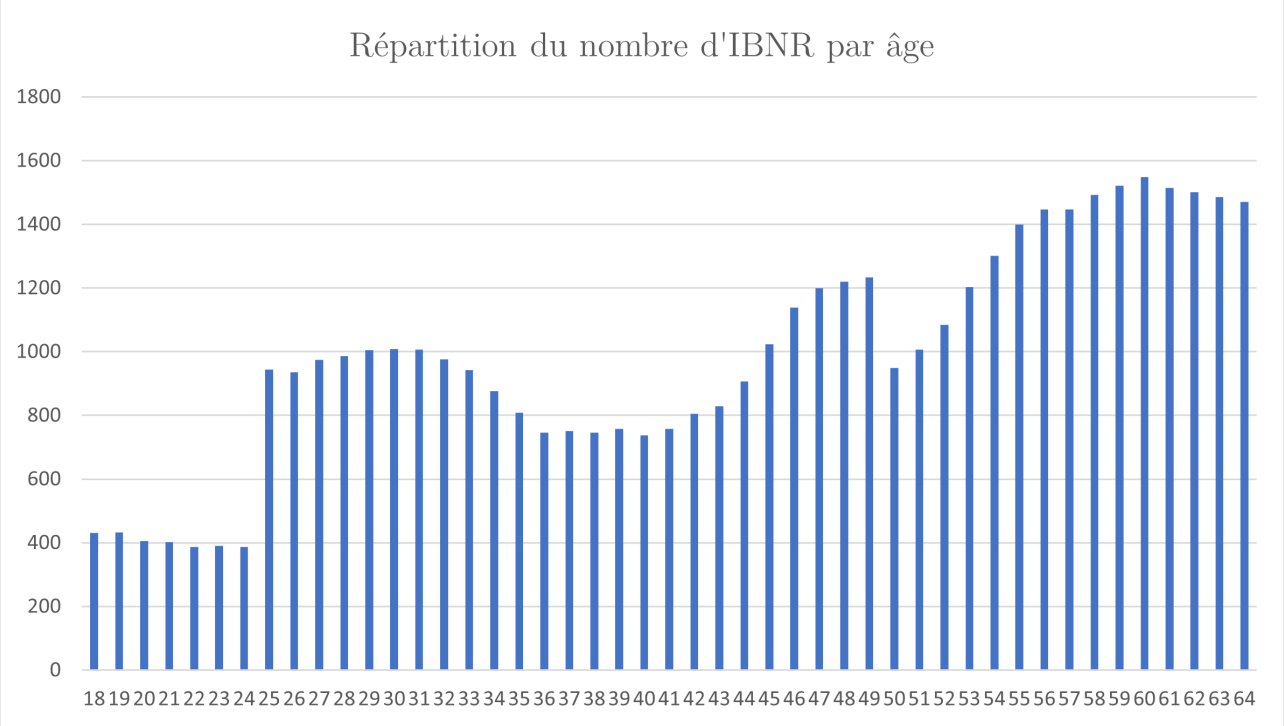
Nombre d'IBNR estimé	Nombre d'IBNR simulé	Ecart en pourcentage
46 517	45 425	2,40%

FIGURE 3.12 : Résultat de l'estimation du nombre d'IBNR

Cette méthode surestime légèrement le nombre d'IBNR.

Afin d'estimer les durées d'indemnisation de ces tardifs, nous nous basons sur une approche ligne à ligne. Nous répartissons le nombre d'IBNR, que nous venons d'estimer, par âge selon la même représentation des entrées en arrêt de travail que nous avons estimée précédemment (figure **??**).

Nous obtenons ainsi une répartition du nombre d'IBNR par âge :



Nous effectuons ensuite une approche âge par âge de la même manière que la méthode de provisionnement ligne à ligne :

Puisque nous n'avons pas d'information sur l'ancienneté des tardifs et puisque le délai de déclaration est d'en moyenne 2 mois, nous supposons que les sinistres tardifs ont une ancienneté de 2 mois.

A partir de cette hypothèse, pour chaque âge, nous calculons le coefficient de provisionnement de la même manière que l'approche ligne à ligne ; nous multiplions ensuite ce coefficient avec le nombre d'IBNR correspondant.

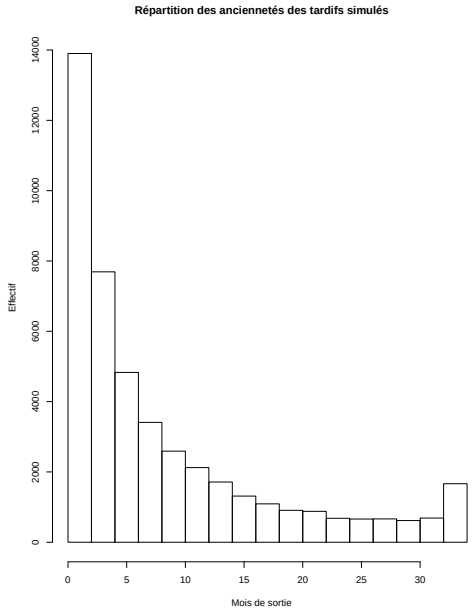
Nous obtenons ainsi une estimation de la charge totale des IBNR :

Charge d'IBNR estimée	**Charge d'IBNR simulée**	**Ecart en pourcentage**
427 764	346 215	23,55%

FIGURE 3.13 : Résultat de l'estimation du nombre d'IBNR

Cette méthode surestime assez fortement la charge d'IBNR. En effet, le nombre d'IBNR estimé semble correct mais la méthode ligne à ligne utilisée pour estimer la durée d'indemnisation de ces tardifs entraine une surestimation importante. L'introduction de l'hypothèse plutôt forte sur l'ancienneté des sinistres tardifs joue un rôle dans cet écart.

En effet si nous regardons la répartition de l'ancienneté des sinistres tardifs simulés :



Nous constatons d'après cet histogramme, que la majorité des sinistres tardifs ont en réalité une ancienneté de 0 et de 1 mois et donc l'hypothèse que nous avons établie semble biaiser l'estimation. Cependant, cette hypothèse était nécessaire à cause du manque d'informations de l'assureur sur les tardifs.

Nous allons maintenant nous intéresser à la distribution empirique des charges estimées et simulées sur un échantillon de 500 simulations.

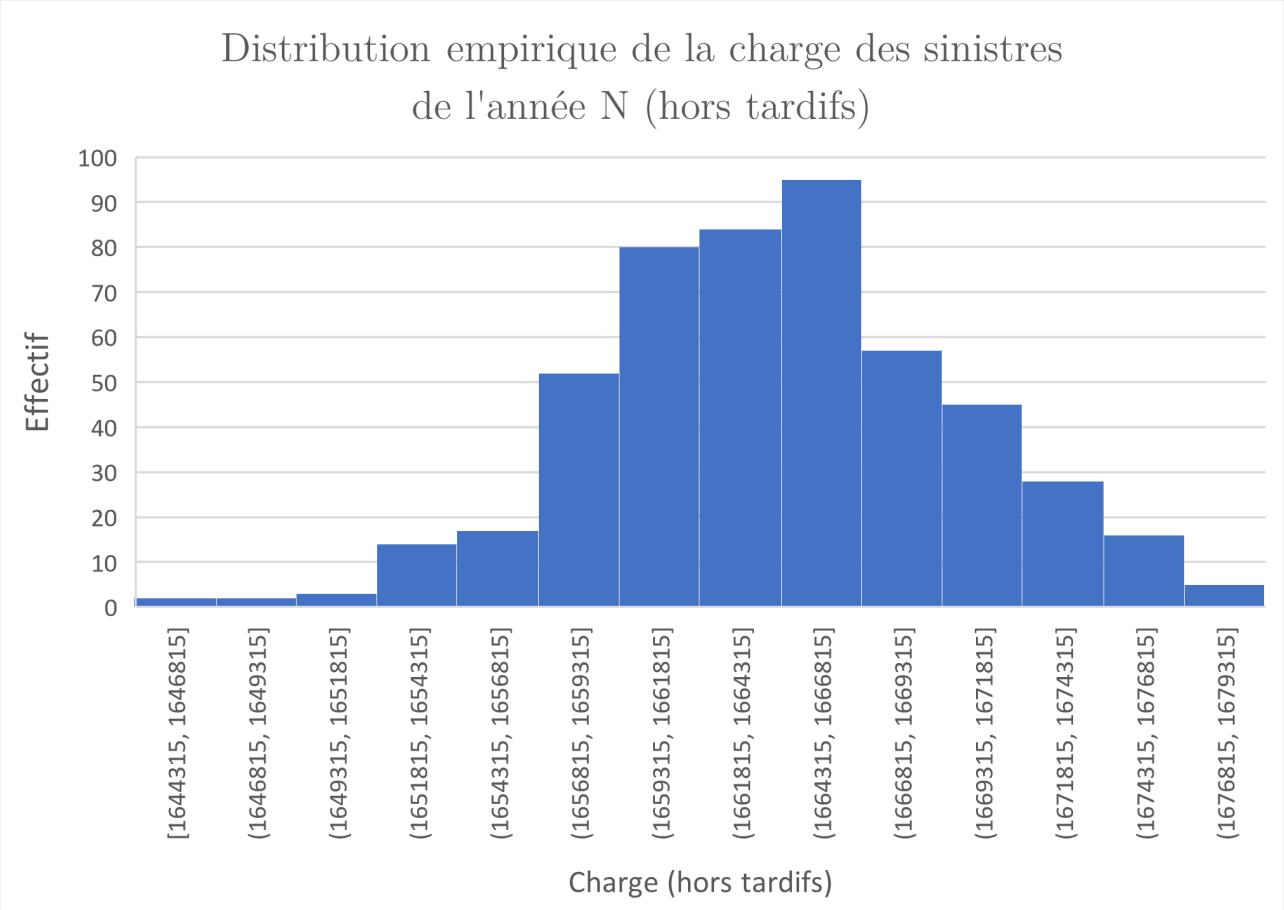
3.5 Résultats des différentes méthodes de provisionnement sur 500 simulations

Le travail de provisionnement a été effectué sur une simulation de la réalité.
Le but de cette section est de répéter le travail présenté dans l'organigramme sur un échantillon de 500 simulations afin d'avoir une représentation de la distribution empirique de la charge ultime de l'année N et de la charge estimée à l'aide des méthodes de provisionnement.
Nous obtenons ainsi leur espérance et leur volatilité empirique que nous pouvons comparer.

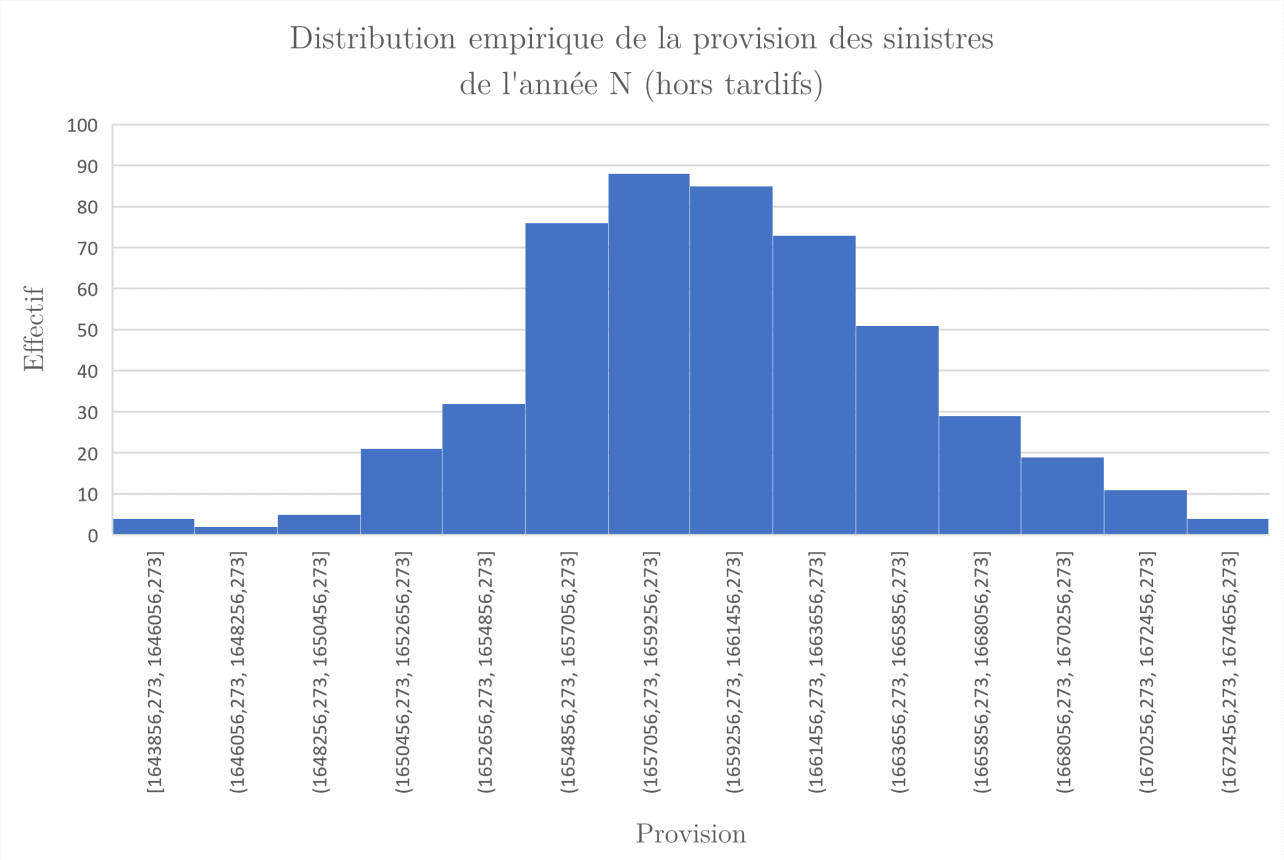
3.5.1 La méthode de provisionnement ligne à ligne

Nous rappelons que cette méthode permet d'estimer la durée d'indemnisation des sinistres en cours. Nous obtenons donc une estimation de la charge des sinistres en cours à laquelle nous ajoutons la charge des sinistres clos. Ce qui constitue une charge de sinistres hors IBNR que nous comparons à la charge simulée hors IBNR.

Sur 500 simulations, nous obtenons la distribution de la charge des durées indemnisées (hors tardifs) suivante :



La distribution de la provision des sinistres à indemniser (hors tardifs) à l'aide de la méthode de provisionnement ligne à ligne sur 500 simulations a l'allure suivante :



Nous obtenons ainsi les espérances et volatilités suivantes :

	Charge estimée (hors IBNR)	Charge simulée (hors IBNR)
Espérance empirique	1 660 071,77	1 664 256,56
Volatilité empirique	5 054,41	5 648,72

FIGURE 3.14 : Résultats de la méthode ligne à ligne sur 500 simulations

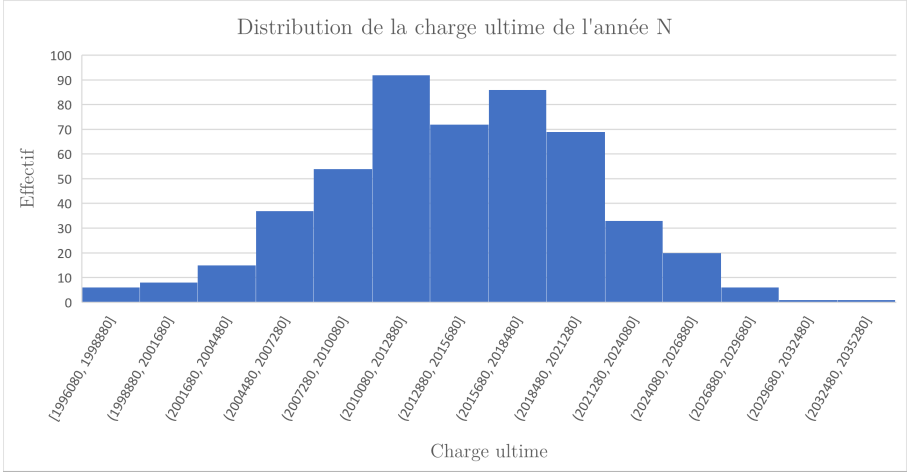
L'espérance empirique de la charge estimée hors IBNR est très proche de la charge simulée avec un écart de $0,25\%$ en valeur absolue, et un écart de $10,5\%$ pour la volatilité empirique.
L'approche ligne à ligne reste donc assez précise et efficace pour provisionner les sinistres en cours d'arrêt de travail.

3.5.2 La méthode de Chain Ladder

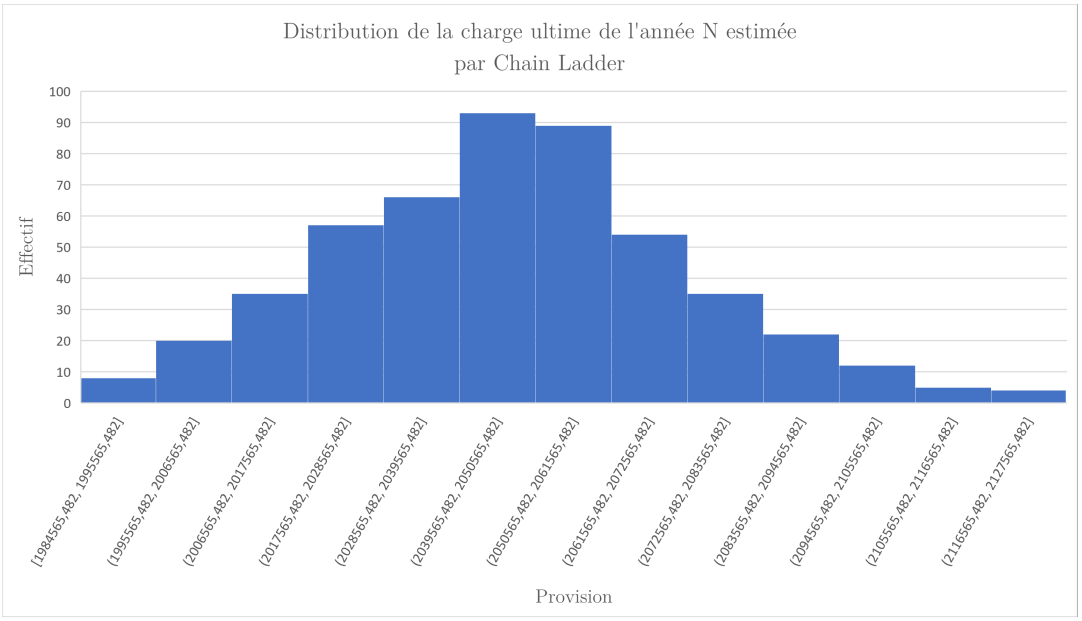
La méthode de Chain Ladder donne une estimation de la charge totale de l'année N en se basant sur un historique de 3 années et donc une estimation des tardifs.

Distribution des charges ultimes

Sur 500 simulations, nous obtenons la distribution de la charge totale de l'année N suivante :



La distribution de la charge ultime estimée à l'aide de la méthode de Chain Ladder sur 500 simulations a l'allure suivante :



Nous obtenons ainsi les espérances et volatilités suivantes :

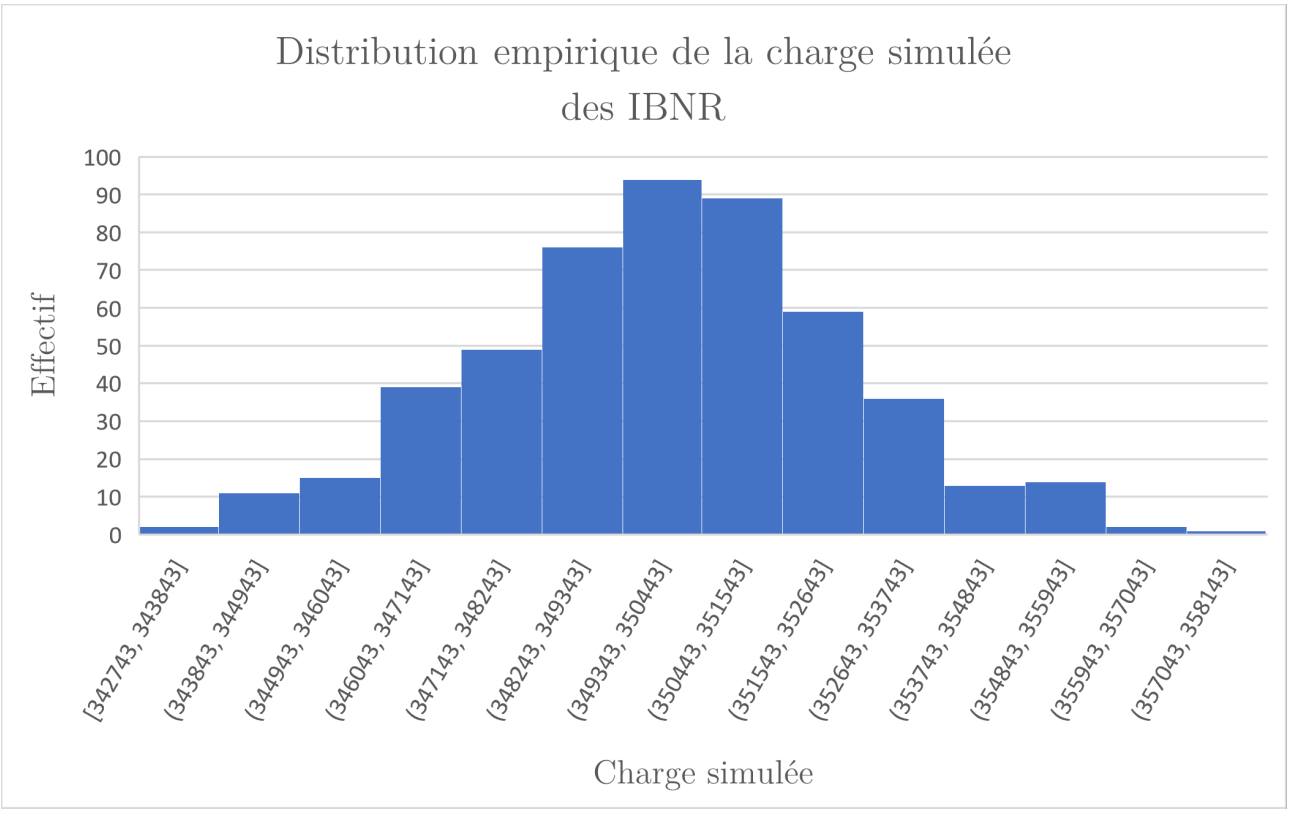
	Charge totale estimée	Charge totale simulée
Espérance empirique	2 047 518,70	2 014 241,00
Volatilité empirique	25 716,05	6 269,44

FIGURE 3.15 : Résultats de la méthode de Chain Ladder sur 500 simulations

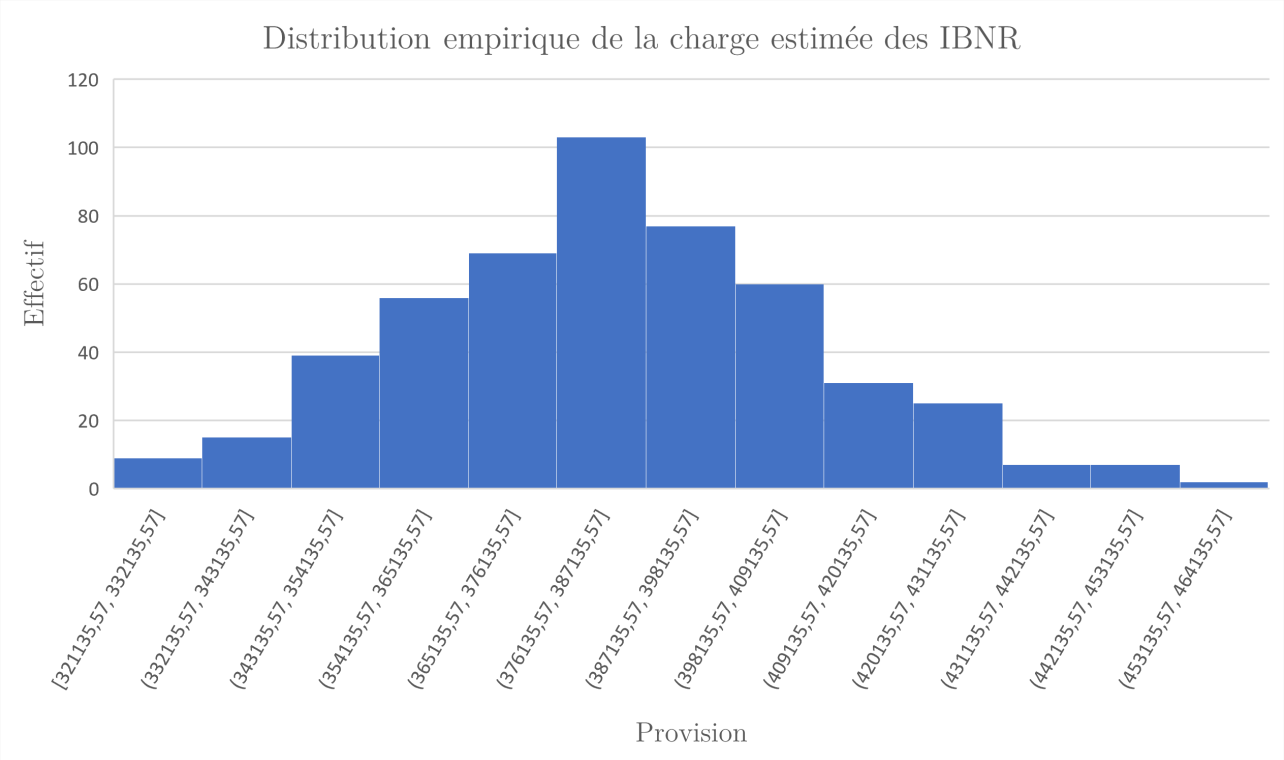
Nous remarquons que l'espérance empirique de la charge totale estimée est assez proche de la charge totale simulée avec un écart de $1,65\%$.
La volatilité empirique est cependant assez importante. Cette méthode reste donc assez efficace pour estimer la charge totale mais demeure très volatile à cause d'une potentielle irrégularité dans l'historique qui peut causer une mauvaise estimation des coefficients de passage.

Distribution des IBNR

Nous obtenons la distribution de la charge des tardifs simulée suivante :



La distribution de la charge des IBNR estimée à l'aide de la méthode de Chain Ladder sur 500 simulations a l'allure suivante :



Nous obtenons ainsi les espérances et volatilités suivantes :

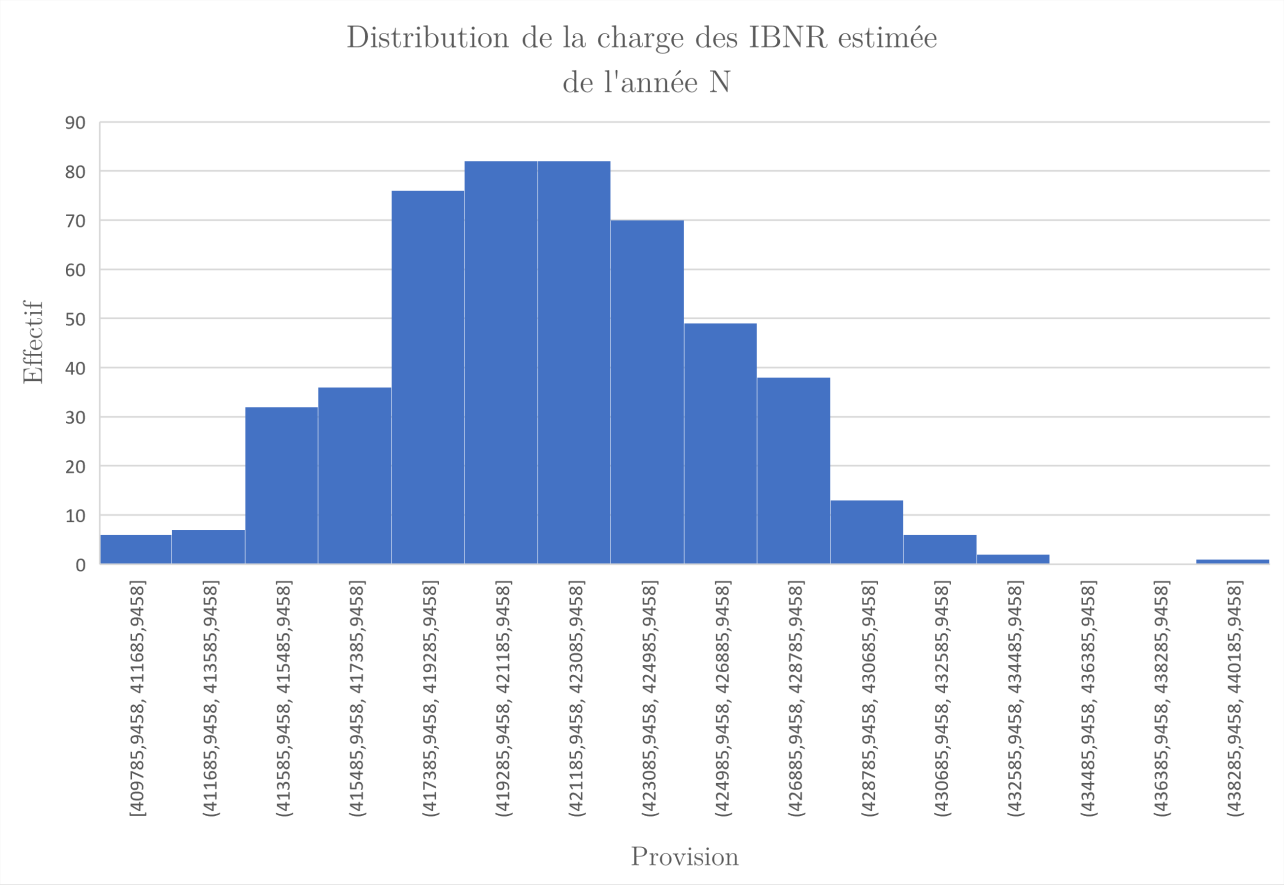
	Charge IBNR estimée	Charge IBNR simulée
Espérance empirique	383 438,40	349 984,43
Volatilité empirique	24 982,65	2 462,10

FIGURE 3.16 : Résultats de la méthode d'estimation des IBNR à l'aide de Chain Ladder sur 500 simulations

L'espérance empirique de la charge des IBNR estimée est supérieure à celle simulée avec un écart de $9,56\%$ et une volatilité assez importante.
La méthode de Chain Ladder surestime la charge des tardifs. Cet écart peut s'expliquer par le choix du calcul du coût moyen. En effet, celui-ci a été calculé sur l'historique total des sinistres clos des 3 années. Il donne donc une estimation assez correcte de la charge totale (section précédente) mais ne représente pas forcément bien le coût spécifique des tardifs.

3.5.3 Méthode de provisionnement à l'aide de la loi d'incidence

Cette méthode a consisté à estimer le nombre d'IBNR à l'aide de la loi d'incidence puis d'effectuer une approche ligne à ligne sur ces derniers en se basant sur une hypothèse assez forte due à un manque d'informations.
Nous obtenons la distribution de la charge des IBNR estimée à l'aide de cette méthode sur 500 simulations :



Nous obtenons ainsi les espérances et volatilités suivantes :

	Charge IBNR estimée	Charge IBNR simulée
Espérance empirique	421 542,37	349 984,43
Volatilité empirique	4 355,39	2 462,10

FIGURE 3.17 : Résultats de la méthode d'estimation des IBNR à l'aide de la loi d'incidence sur 500 simulations

L'espérance empirique de l'estimation de la charge des IBNR avec cette méthode est assez importante comparée à la charge simulée avec un écart de $20,45\%$. Cet écart s'explique par l'hypothèse sur l'ancienneté des individus supposée de 2 mois qui est assez forte et cela entraine donc un surprovisionnement lors du calcul du coefficient de provisionnement.

Cette hypothèse semblait cependant nécessaire à prendre en compte quant au manque d'informations sur les tardifs. Elle est également de bon sens puisque le délai de déclaration des sinistres est d'en moyenne 2 mois.

3.5.4 Comparaison des différentes méthodes de provisionnement

Dans cette partie, nous confrontons les résultats de la section précédente sur l'estimation et la simulation en fonction des différentes méthodes.
La méthode de provisionnement ligne à ligne donne une estimation des sinistres en cours de l'année N. Cette estimation ajoutée à la charge des sinistres clos donne une estimation de la charge des sinistres de l'année N hors tardifs.
Nous avons vu que cette méthode était assez précise et peu volatile avec un écart de $0,25\%$ avec la simulation.

La méthode de Chain Ladder quant à elle, en se basant sur un historique de 3 ans, donne une estimation de la charge ultime des sinistres de l'année N y compris des IBNR. Cette méthode surprovisionne la charge totale de l'année N avec un écart de $1,65\%$. De plus cette méthode reste très volatile due à l'irrégularité de l'historique.

Ces deux méthodes sont les plus fréquemment utilisées par les assureurs pour provisionner les sinistres d'arrêt de travail. Elles possèdent chacune leurs avantages et leur limites.

Avantages des méthodes ligne à ligne et agrégées

Les méthodes ligne à ligne sont des méthodes assez précises car le risque est individualisé en fonction des caractéristiques de chaque assuré. Elles ont l'avantage d'identifier les différents facteurs associés au risque d'arrêt de travail. Elles permettent donc d'approcher de manière assez fine les prestations futures.

Les méthodes agrégées quant à elles ne nécessitent pas de caractéristiques individuelles. Lors de la mise en place de ces méthodes, les sinistres en cours et les tardifs sont globalisés et donc cela ne nécessite pas une approche spécifique pour chacun. Elles sont finalement plus simples à mettre en place et faciles à interpréter.

Limites des méthodes ligne à ligne et agrégées

Lors d'une approche ligne à ligne, les assureurs ont besoin d'un portefeuille avec une quantité d'informations beaucoup plus importante au niveau individuel. Elles peuvent également nécessiter la construction de table d'expérience avec des hypothèses pertinentes pour refléter au mieux la population assurée. Leur mise en oeuvre peut être donc relativement complexe et contraignante.

Les méthodes agrégées quant à elles sont moins précises, assez volatiles et entrainent souvent un surprovisionnement et une surestimation du risque comme nous l'avons vu. Elles nécessitent également une structure du portefeuille assez homogène d'année en année et donc un historique assez important afin de les mettre en place.

Nous allons maintenant comparer les méthodes mises en oeuvre pour l'évaluation et le provisionnement spécifique des tardifs.

Comparaison des méthodes utilisées pour évaluer les tardifs

Dans ce mémoire, le provisionnement des tardifs a nécessité une évaluation à part entière selon deux méthodes :

- Nous avons réalisé du Chain Ladder sur un triangle en nombre de sinistres afin d'estimer le nombre d'IBNR que nous avons multiplié par un coût moyen choisi en fonction de l'historique.
- Par ailleurs, nous avons utilisé la loi d'incidence de Prim'Act pour estimer le nombre d'IBNR. Ensuite, nous avons réalisé une approche ligne à ligne sur ces derniers à l'aide d'une hypothèse sur l'ancienneté de 2 mois des sinistres tardifs.

Nous obtenons les résultats synthétiques suivants sur 500 simulations :

	Charge des sinistres tardifs
Espérance empirique de la simulation	349 984,43
Charge des tardifs estimée avec Chain Ladder	
Espérance empirique	383 438,40
Volatilité empirique	24 982,65
Ecart relatif à la simulation	9,56%
Charge des tardifs estimée avec la loi d'incidence	
Espérance empirique	421 542,37
Volatilité empirique	4 355,39
Ecart relatif à la simulation	20,45%

FIGURE 3.18 : Résultats des méthodes sur l'évaluation des IBNR

Les deux méthodes surestiment la charge des IBNR. La méthode de Chain Ladder entraine un surprovisionnement des sinistres tardifs avec un écart de $9,5\%$ contre un écart de $20,45\%$ avec l'autre méthode.

La méthode de Chain Ladder reste plus précise que l'autre méthode et l'écart de $9,5\%$ peut s'expliquer d'une part à cause d'une surparametrisation du modèle qui entraine un nombre d'IBNR plus important que la simulation et d'autre part à cause du choix du coût moyen, qui certes représente bien l'historique complet, mais moins bien la part des sinistres tardifs.
L'écart de la méthode basée sur la loi d'incidence quant à lui s'explique par l'hypothèse sur l'ancienneté.

La méthode de Chain Ladder reste cependant beaucoup plus volatile que la méthode basée sur la loi d'incidence qui elle a nécessité l'intervention d'une approche ligne à ligne.

Pour conclure, la méthode de Chain Ladder est plus précise que la méthode basée sur la loi d'incidence pour estimer les tardifs mais reste cependant beaucoup plus volatile.

En ce qui concerne la méthode basée sur la loi d'incidence, il aurait peut être fallu complexifier l'hypothèse sur l'ancienneté des sinistres tardifs afin d'améliorer la précision de l'estimation. Par exemple, on aurait pu reprendre la représentation de l'ancienneté des sinistres clos de l'année N et supposer que les tardifs ont la même structure, puis utiliser la méthode ligne à ligne sur cette hypothèse.

Cela aurait peut être amelioré l'écart avec la simulation tout en gardant une volatilité assez faible grâce à l'approche ligne à ligne implémentée. Cette méthode aurait donc peut être surpassé la méthode de Chain Ladder en terme de précision mais aurait complexifié les calculs.

Chapitre 4

Etude de la stabilité des méthodes de provisionnement suite à une erreur de modélisation et une connaissance imparfaite du portefeuille assuré

L'objectif de ce chapitre est de regarder les résultats obtenus d'une part lorsqu'il y'a une erreur d'estimation dans la construction des tables d'entrée et de maintien utilisées pour provisionner et d'autre part lorsque le portefeuille considéré possède une structure imparfaite.

4.1 Impact de l'erreur d'estimation des tables sur le provisionnement

Nous nous plaçons dans le cas où l'assureur construit ses propres tables afin de provisionner les sinistres. La table d'entrée en incapacité est utilisée pour estimer le nombre de sinistres tardifs et la table de maintien est utilisée pour estimer la charge des sinistres en cours. Dans cette section, nous considérons l'erreur d'estimation des tables suite à leur construction et leur impact sur le provisionnement.

Afin de traduire l'erreur d'estimation, nous allons introduire différents chocs sur les taux q_x^{inc} de la table d'incidence de Prim'Act et sur les taux q_x^{main} de la table de maintien du BCAC puis regarder les résultats de l'estimation.

4.1.1 Impact des chocs de la table d'incidence sur l'estimation

La table d'incidence est utilisée afin d'estimer le nombre d'entrées en incapacité qui permet de déduire le nombre de sinistres tardifs dans l'année.
Nous allons regarder comment évolue l'écart de l'estimation du nombre de tardifs avec la simulation en fonction de l'erreur d'estimation de la table d'incidence.

Chocs uniformes

Les taux q_x^{inc} de la table d'incidence sont choqués de manière uniforme pour chaque âge. Lorsque nous estimons les tardifs à l'aide de cette nouvelle table choquée, nous obtenons les résultats suivants en fonction des différents chocs :

Chocs	Nombre d'IBNR estimé	Nombre d'IBNR simulé	Ecart relatif
10,00%	72 901	45 425	60,49%
5,00%	59 709	45 425	31,45%
2,00%	51 792	45 425	14,02%
1,00%	49 156	45 425	8,21%
0,50%	47 841	45 425	5,32%

Différents chocs ont été établis sur la table d'incidence et on peut observer leur impact sur l'écart relatif. On remarque qu'un choc de 10% entraine un écart relatif assez important de 60.49%. Les chocs de 0.50% à 5.00% entrainent des écarts moins importants.
Dans la partie suivante, nous discuterons de la validité des ordres de grandeurs des chocs en les comparant avec ce que l'on peut trouver en pratique.

Chocs par tranches d'âge

Nous avons également choqué la table en fonction des tranches d'âge. 5% pour les tranches jeunes de 18 à 25 ans, 3% pour les tranches de 25 à 35 ans, 2% pour les tranches de 35 à 50 ans et 5% pour les tranches de 50 à 64 ans.
Nous avons affecté un choc plus important pour les âges extrêmes car l'effectif du portefeuille étant moins important sur ces tranches d'âge, cela peut entrainer une erreur plus importante lors de la construction.

Nous obtenons les résultats suivants :

Nombre d'IBNR estimé	Nombre d'IBNR simulé	Ecart relatif
56 023	45 425	23,33%

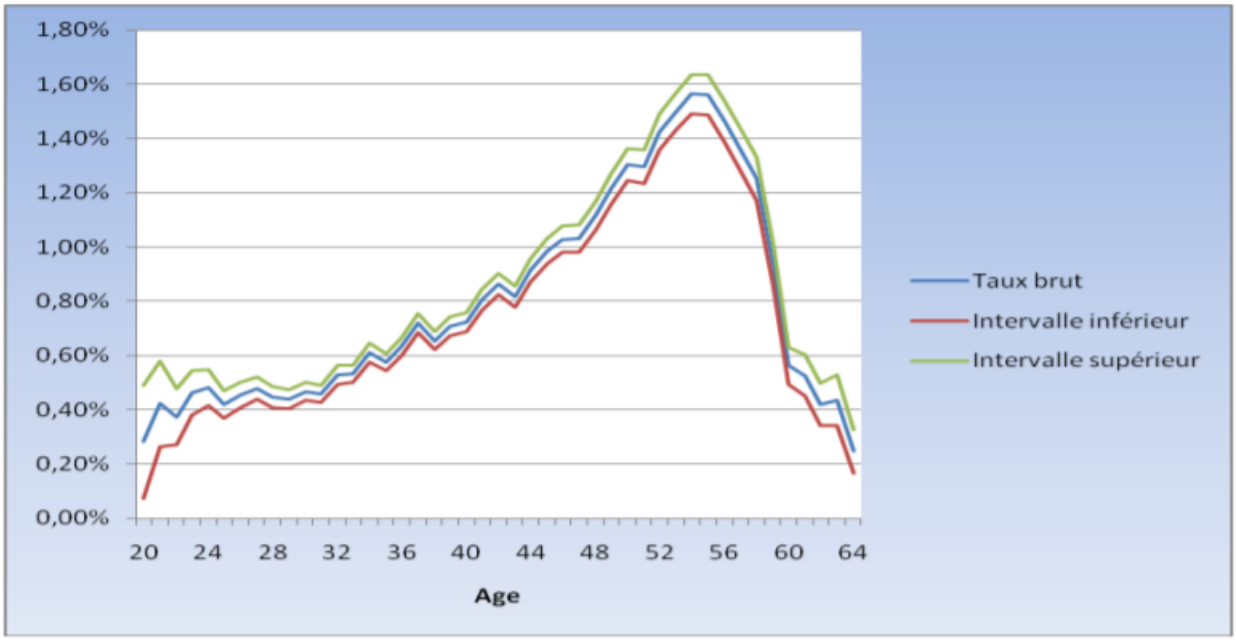
Pertinence du choix des chocs sur la table d'incidence

La table d'incidence de Prim'Act est construite pour des contrats d'assurance emprunteur. Nous allons regarder en pratique l'ordre de grandeur de l'erreur d'estimation lors de la construction de table d'entrées en arrêt de travail.
Le mémoire de Tom LEURENT "Construction de tables d'expérience des risques incapacité et invalidité" traite de la construction d'une table d'entrée en incapacité à partir d'un portefeuille d'assurance emprunteur.

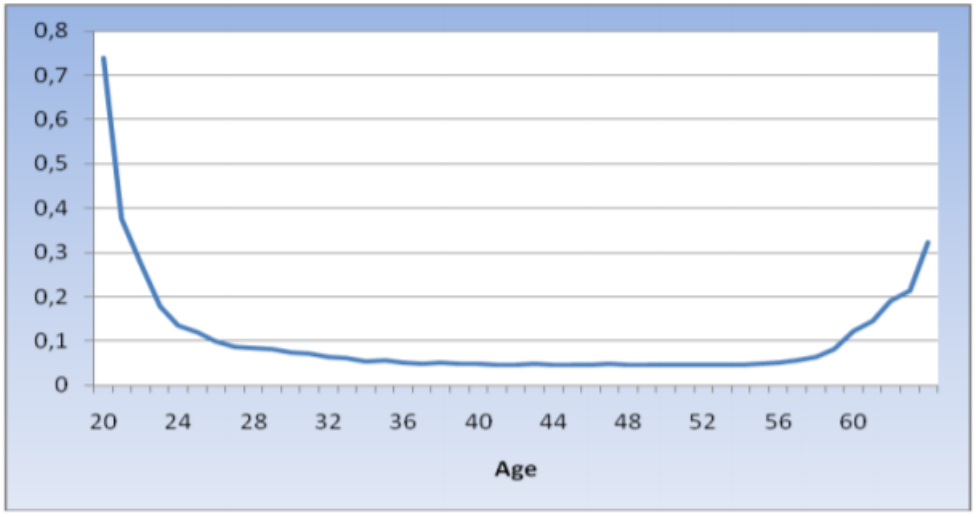
Le portefeuille emprunteur étudié possède un historique de 5 ans avec à peu près 1 million de contrats observés. Nous regardons les intervalles de confiance des résultats sur les taux bruts ainsi obtenus :

Année	Nombre de contrats observés	Nombre de personnes observées	Evolution
2004	1 025 413	1 287 571	
2005	1 071 622	1 339 390	+ 4,51%
2006	1 132 269	1 419 226	+ 5,66%
2007	1 156 767	1 454 076	+ 2,16%
2008	1 171 286	1 470 958	+ 1,26%
Ensemble	1 783 273	2 262 575	

Tableau 4 : Nombre de contrats et de personnes observés sur le portefeuille emprunteur



Graphique 18 : intervalles de confiance des taux bruts d'entrée en incapacité du portefeuille emprunteur



Graphique 19 : Rapport entre les intervalles de confiance et les taux bruts

Le dernier graphique mesure l'écart entre le taux brut et la borne de l'intervalle de confiance en valeur absolue. Nous remarquons que le rapport est important pour les âges extrêmes et reste constant en moyenne à 0.5 pour la tranche de 30 à 56 ans.
L'ordre de grandeur des chocs proposés semble donc correct.

4.1.2 Impact des chocs de la table de maintien sur l'estimation

La table de maintien est utilisée pour provisionner les sinistres en cours. Nous affectons différents chocs de manière uniforme pour chaque âge et regardons l'impact de ces chocs sur l'estimation de la charge des sinistres en cours.

Chocs uniformes

Les taux de la table de maintien sont choqués de manière uniforme pour chaque âge. Lorsque nous estimons la charge des sinistres en cours à l'aide de la nouvelle table, nous obtenons les résultats suivant en fonction des chocs :

Chocs	Charge estimée	Charge simulée	Ecart relatif
20,00%	1 817 672	1 656 717	9,72%
15,00%	1 776 112	1 656 717	7,21%
10,00%	1 734 553	1 656 717	4,70%
5,00%	1 692 994	1 656 717	2,19%

Nous observons que l'écart varie peu en fonction des chocs à la différence de la table d'incidence.
Nous effectuons également un choc en fonction de l'âge à la survenance par tranche d'âge de la même manière que la table d'incidence.

Chocs par tranches d'âge

Nous avons choqué la table de maintien par tranches d'âge de la manière suivante : 20% pour les tranches jeunes de 18 à 25 ans, 15% pour les tranches de 25 à 35 ans, 10% pour les tranches de 35 à 40 ans, 15% pour les tranches de 41 à 50 ans et 20%
Nous avons affecté un choc plus important pour les tranches aux âges extrêmes car l'effectif du portefeuille étant moins important pour ces tranches, cela peut entrainer une erreur plus importante lors de la construction.
Nous obtenons les résultats suivants :

Charge estimée	Charge simulée	Ecart relatif
1 795 007	1 656 717	8,35%

Pertinence du choix des chocs sur la table de maintien

Nous allons regarder en pratique l'ordre de grandeur de l'erreur d'estimation lors de la construction de table de maintien en arrêt de travail.
Différents mémoires traitent de la construction de table de maintien.
Le mémoire de FINAS Céline sur "Modélisation du risque de maintien en incapacité" traite du sujet. L'étude a été faite sur une base de données de sinistres du portefeuille d'Axa survenus entre 1996 et 2008.
Les résultats sur la largeur de l'intervalle de confiance des taux bruts obtenus pour différentes tranches d'âge sont les suivants :

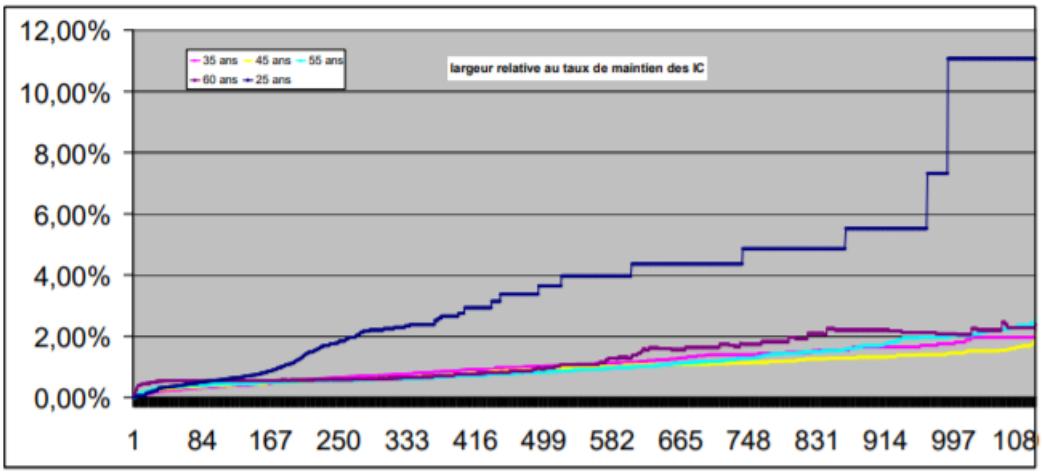


Figure 19 largeur relative de l'intervalle de confiance des taux bruts obtenus en fonction de l'ancienneté du sinistre

La largeur de l'intervalle de confiance varie de 0 à 11% pour les sinistres à 25 ans tandis la largeur est autour de 2% pour les autres âges.

Le mémoire de Ludovic CAMOL sur "Provisionnement de la garantie Incapacité/Invalidité Construction d'une table d'expérience" traite également du sujet. L'étude est faite sur une base de données d'un portefeuille de 2013 contenant 112 616 sinistres après traitement.

La largeur relative de l'intervalle de confiance pour l'âge de 54 ans obtenue est la suivante :

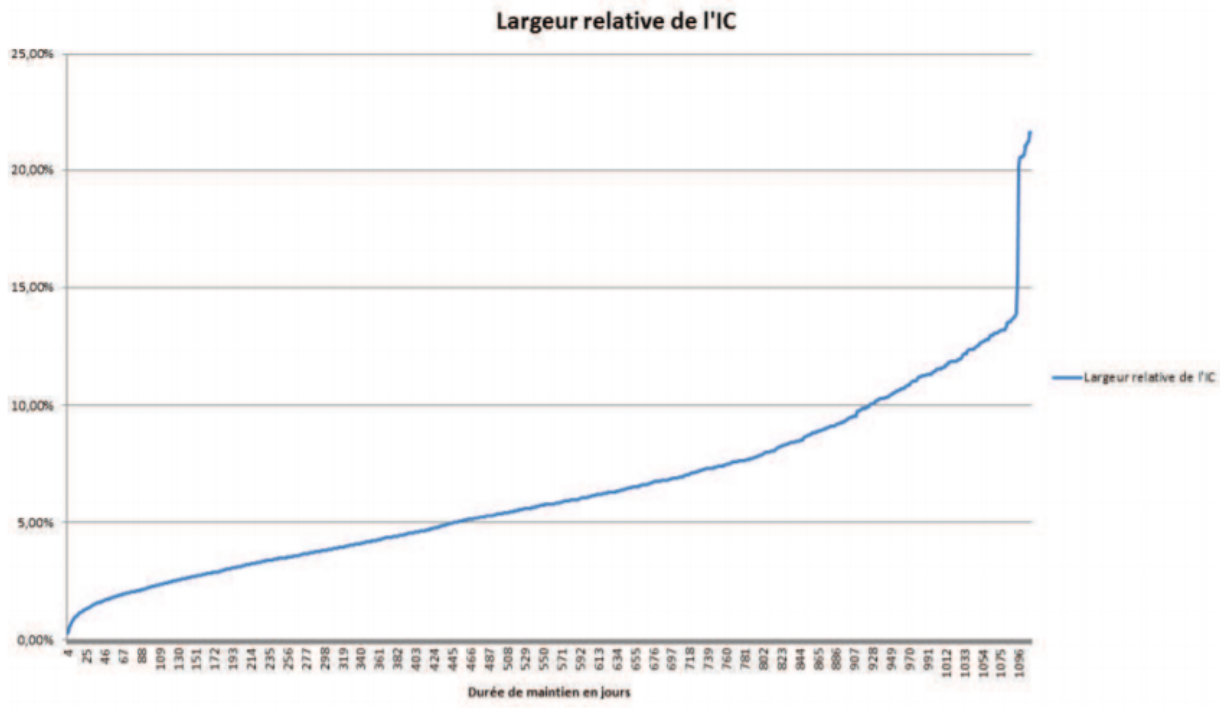


FIGURE 40 – Largeur relative de l'intervalle de confiance pour l'âge 54 ans

La largeur varie de 0 à 20% en fonction de l'ancienneté.

On peut conclure que les chocs proposés de 5%, 10% et 15% en ce qui concerne la table de maintien sont des ordres de grandeur cohérents.
Nous allons maintenant étudier le cas où l'assureur manque d'information sur sa base de sinistres.

4.2 Impact d'une structure imparfaite du portefeuille sur le provisionnement

Dans cette partie nous étudions l'impact d'une structure imparfaite du portefeuille sur le provisionnement ligne à ligne.

Nous nous plaçons dans le cas où l'assureur ne reçoit pas une base ligne à ligne avec les informations complètes concernant chaque individu.
L'assureur reçoit une base pour l'année N comportant les variables suivantes : les tranches d'âge à la survenance, l'effectif des individus sinistrés par tranche d'âge ainsi que l'ancienneté moyenne des individus par tranches d'âge.

Afin de provisionner cette base de sinistres en cours, l'assureur a besoin de faire des hypothèses à cause du manque d'informations.
Pour chaque tranche d'âge, nous considérons un individu moyen avec deux caractéristiques : âge à la survenance et ancienneté (ces caractéristiques sont choisies en fonction de l'historique du portefeuille et leur choix sera expliqué dans la suite). Nous estimons ensuite la durée de maintien des individus à l'aide du coefficient de provisionnement ligne à ligne calculé à partir des caractéristiques de l'individu moyen. Ce coefficient est ensuite multiplié par l'effectif de la tranche d'âge correspondante afin d'obtenir la charge totale estimée des sinistres en cours.

Hypothèses sur l'individu moyen

Afin de provisionner, nous considérons un individu moyen pour chaque tranche d'âge. Les caractéristiques de l'individu moyen, âge et ancienneté sont choisies de la manière suivante :

- Pour l'âge à la survenance on choisit l'âge moyen de la tranche d'âge à la survenance sur l'historique des années $N-2$ et $N-1$. Par exemple pour la tranche d'âge 18 à 25 ans, la moyenne d'âge des sinistres de l'historique est de 22 ans. On choisit donc 22 pour l'âge de l'individu moyen de cette tranche. ;
- L'ancienneté de l'individu moyen correspond à l'ancienneté moyenne donnée de la base.

On dispose ainsi d'une base comportant les individus moyens qui correspondent à chaque tranche d'âge.

Provisionnement et résultats

Nous estimons les durées de maintien suivant la méthode ligne à ligne à l'aide de la table du BCAC en calculant le coefficient de provisionnement à l'aide des caractéristiques de chaque individu moyen. Nous multiplions ensuite ces coefficients aux effectifs des tranches d'âges et nous obtenons ainsi l'estimation de la charge des sinistres de l'année N de la base.

Charge estimée	Charge simulée	Ecart relatif
1 711 749	1 656 717	3,32%

Nous observons que l'écart est beaucoup plus important lorsque le portefeuille manque d'informations individuelles (3.32% contre 0.25% lorsque le portefeuille est complet) mais reste correct.

Nous regardons maintenant les résultats lorsque nous utilisons les tables de maintien choquées de la partie précédente pour estimer les charges.

Résultats du provisionnement avec les tables d'expérience

Chocs	Charge estimée	Charge simulée	Ecart relatif
5,00%	1 756 324	1 656 717	6,01%
10,00%	1 800 899	1 656 717	8,70%
15,,00%	1 845 474	1 656 717	11,39%

Nous observons que lorsque nous rajoutons l'incertitude liée à la table d'expérience l'écart augmente significativement (quasiment le double pour un choc de 10%).

4.3 Conclusion sur la stabilité des méthodes

Ce chapitre a eu pour objectif d'observer l'écart entre l'estimation et la simulation lorsque l'erreur provient de la modélisation (construction des tables de l'assureur en fonction de ses données) ainsi que lorsque le portefeuille de l'assuré présente un manque d'informations.

L'estimation du nombre d'IBNR à l'aide de la table d'entrée en incapacité peut entrainer un biais important en fonction de l'incertitude de la table. De plus, l'estimation des sinistres en cours à l'aide de la table de maintien nécessite une quantité d'information individuelle importante. Lorsqu'une partie des informations sont manquantes, cela oblige l'assureur à utiliser des hypothèses ce qui introduit un biais dans l'estimation.

Les méthodes de provisionnement agrégées comme Chain Ladder sont plus robustes. En effet, elles ne nécessitent pas de caractéristiques individuelles et ne sont donc pas impactées par le manque d'information que peut contenir un portefeuille d'assurance. Il est donc plus intéressant d'utiliser des méthodes agrégées dans ce cas-là.

Conclusion

Le provisionnement est un enjeu crucial pour l'assureur qui cherche à prédire les prestations qu'il aura à verser dans le futur. La part des provisions est un élément important du passif de l'assurance comme nous l'avons vu.

L'objectif de ce mémoire était de confronter les différentes méthodes de provisionnement des sinistres en cours et des tardifs, utilisées dans le marché dans le cadre de l'arrêt de travail avec une réalité simulée de toute la population active française.

Nous avons donc réalisé une simulation à l'aide de données et d'hypothèses spécifiques afin de représenter au mieux et de manière simple le comportement et la sinistralité d'une population active importante.

A ce stade, nous nous sommes placés du point de vue de l'assureur au $31/12/N$ qui cherche à provisionner les sinistres en cours et les tardifs de son portefeuille. Traditionnement, les méthodes utilisées sont les méthodes ligne à ligne et les méthodes agrégées qui ont nécessité la simulation d'un historique sur 3 ans.

Au début, nous avons utilisé ces méthodes sur un état du monde puis sur 500 simulations afin d'obtenir une distribution empirique de l'estimation.

Les résultats obtenus, confrontés à la réalité simulée, avec l'approche ligne à ligne ont été plutôt concluants et meilleurs que les résultats de la méthode de Chain Ladder. Cependant, l'estimation des tardifs par la méthode agrégée a été plus précise mais aussi plus volatile que la méthode basée sur la loi d'incidence. En effet, celle-ci a nécessité l'introduction d'une hypothèse assez forte due à un manque d'informations sur les tardifs.

Il aurait donc pu être intéressant de compléter l'étude en introduisant une hypothèse plus complexe sur l'ancienneté des sinistres tardifs puis de recomparer les deux méthodes entre elles.

Bien qu'il soit intéressant de changer la méthode d'estimation, nous avons également vu dans la dernière partie les limites de la modélisation liées à une erreur d'estimation lors de la construction des tables d'expérience. En effet, les méthodes de provisionnement ligne à ligne nécessitent la construction de tables à partir des données du portefeuille de l'assureur. Cette construction étant imparfaite, cela peut entrainer un biais lors du provisionnement comme nous l'avons vu.
De plus, les méthodes de provisionnement ligne à ligne ont besoin d'informations individuelles. Lorsqu'il y'a un manque de données au sein du portefeuille, cela amplifie l'incertitude de l'estimation de la provision. Les méthodes agrégées quant à elles supportent beaucoup plus ce phénomène, elles sont plus robustes et donc plus efficaces dans ce contexte.

BIBLIOGRAPHIE

Cours

DUTANG C. [2019-2020], *Actuariat de l'Assurance Non-Vie*

GUIBERT Q. [2019-2020] , *Modèles de simulation en assurance*

JUILLARD M. [2010-2011] , *Gestion et Mesure des risques*

LE VALLOIS F. [2019-2020], *Réglementation prudentielle en assurance*

LOPEZ O.[2019-2020], *Modèles de durée*

Articles

PLANCHET F., JACQUEMIN J.[2003], *L'Utilisation de méthodes de simulation en assurance*

PLANCHET F., THEROND P. [2006], *Modèles de durée Applications actuarielles*

Mémoires

CAMOL L.[2014], *Provisionnement de la garantie Incapacité/Invalidité Construction d'une table d'expérience*

DE LA MORINERIE A. [2016], *Conceptualisation d'un modèle multi-états en arrêt de travail et application à une loi d'incidence en incapacité*

FINAS C.[2009], *Modélisation du risque de maintien en incapacité*

LEURENT T.[2010], *Construction de tables d'expérience des risques incapacité et invalidité*

LOT S.[2019], *Définition des critères de choix du modèle de provisionnement et illustration avec des données d'arrêt de travail*

NGOC AN DINH & GILLES CHAU [2012], *Mesures de provision cohérentes et méthodes ligne à ligne pour des risques non-vie*

Publications et rapports

Fédération Française de l'Assurance [2019], *Rapport 2019*

INSEE [2019]- *Tableaux de l'économie française* Edition 2019

INSEE [2019] - *Bilan démographique 2019*

Optimind Les dossiers techniques d'information Optimind [2010], *La Prévoyance*